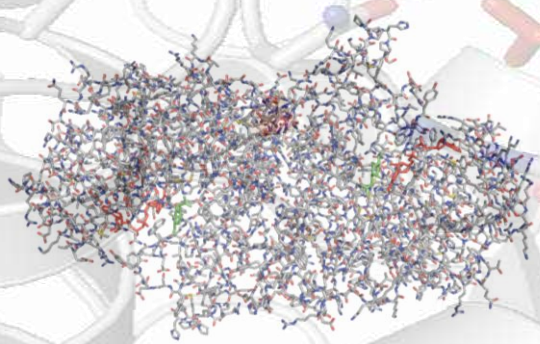
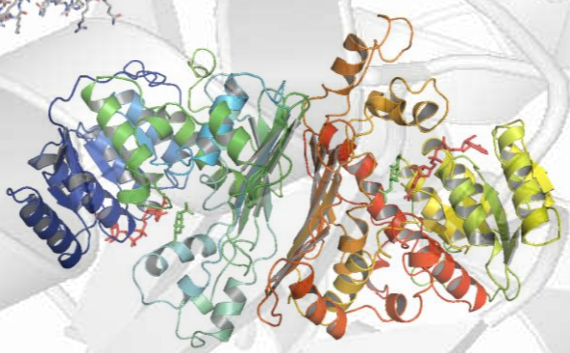
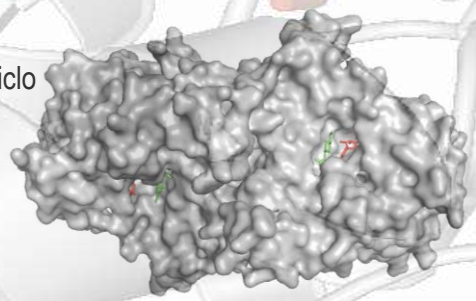
Enzymes Discovery and Characterization: Microbial Hydrolases for Sustainable Bioconversion and Novel Applications

Nicola Curci







Dottorato in Biotecnologie – XXXIII ciclo

Università di Napoli Federico II

Dottorato in Biotecnologie – XXXIII ciclo

Università di Napoli Federico II

ENZYMES DISCOVERY AND CHARACTERIZATION: MICROBIAL HYDROLASES FOR SUSTAINABLE BIOCONVERSION AND NOVEL APPLICATIONS

Nicola Curci

Dottorando:	Nicola Curci
Relatore:	Prof. Marco Moracci
Correlatore	Dr Beatrice Cobucci-Ponzano
Coordinatore:	Prof. Marco Moracci

Settore Scientifico Disciplinare BIO/10

“The practice of any area of research ultimately reduce to a person hands and mind focused on simple objective organized according to a coherent strategy”

Alexander McPherson – Protein Crystallographer

INDICE

RIASSUNTO

Gli enzimi sono biocatalizzatori che accelerano le reazioni biochimiche alla base del metabolismo di tutti gli esseri viventi, catalizzando reazioni con alta specificità e processività. Nel corso degli anni, queste macromolecole hanno avuto un grande impatto nello sviluppo di bioprocessi industriali, consentendo di aumentare le rese e la qualità dei prodotti e riducendo l'utilizzo di energia e l'impatto ambientale.
Il crescente sviluppo di processi ecosostenibili e la necessità di un passaggio verso una bioeconomia circolare ha aumentato la domanda di nuovi biocatalizzatori. Il mercato globale degli enzimi, infatti, è in continuo aumento, con un indice di crescita annuale previsto del 7.2 % tra il 2020 ed 2027. Tale mercato è dominato dalle idrolasi, quali proteasi, idrolasi dei carboidrati ed enzimi lipolitici. Molte industrie operanti in diversi settori (i.e., alimentare, tessile, farmaceutico, produzione di biocarburanti, carta e detergenti) beneficiano dell'uso di questi biocatalizzatori. Tuttavia, la limitata disponibilità commerciale di alcune particolari classi enzimatiche e le scarse prestazioni di molti enzimi nelle tipiche condizioni dei processi industriali (i.e., alte o basse temperature, pH acidi o alcalini, alte concentrazioni di substrato, presenza di solventi organici), rallentano l'introduzione di nuovi biocatalizzatori su larga scala. In questo scenario, l'identificazione di nuovi enzimi con un'elevata attività, stabilità e una maggiore specificità, rispetto i biocatalizzatori già utilizzati, diventa indispensabile; ponendo le basi per migliorare i bioprocessi preesistenti e lo sviluppo di nuove biotecnologie. La ricerca di nuovi enzimi per applicazioni industriali è principalmente indirizzata verso fonti microbiche. Infatti, grazie alla loro abbondanza naturale, la grande biodiversità e la facilità di coltura in condizioni di laboratorio, i microrganismi sono la fonte principale degli enzimi commerciali attualmente utilizzati. Inoltre, rispetto gli enzimi da animali e vegetali, i biocatalizzatori microbici hanno diversi vantaggi: sono più attivi e stabili, vengo prodotti con alte rese e la ricerca di nuovi enzimi dai microrganismi è facilitata dalle moderne tecniche di metagenomica. Quest'approccio, grazie all'evolversi delle tecnologie di NGS, permette l'identificazione di un consistente numero di nuove sequenze codificanti per enzimi direttamente da campioni ambientali, senza la necessità dell'isolamento dei microrganismi. Inoltre, con il suo avvento, è stato facilitato lo studio di consorzi microbici poco esplorati, come i microrganismi estremofili che, da un punto di vista biotecnologico, rappresentano una fonte di enzimi molto stabili e naturalmente adatti alle rigide condizioni dei processi industriali. Tuttavia, l'identificazione massiva di nuove sequenze ha generato una sostanziale discrepanza tra le sequenze enzimatiche depositate e gli

enzimi caratterizzati. La caratterizzazione enzimatica, infatti, rappresenta l'unico metodo per validare le potenzialità di un enzima come biocatalizzatore e l'enorme *gap* generato limita la scoperta di nuove attività interessanti e potenzialmente sfruttabili. In questo contesto, approcci di tipo *genome mining*, interamente dipendenti da strategie computazionali, permettono la deduzione della possibile funzione enzimatica sulla base dalle informazioni ottenute dalla sequenza. Di conseguenza, queste strategie possono aiutare a filtrare e razionalizzare le informazioni contenute nei genomi microbici annotati e nei *dataset* metagenomici, semplificando la selezione di sequenze enzimatiche interessanti ed utili per scopi biotecnologici ed industriali. Tra i biocatalizzatori più ricercati per questi scopi, vi sono le glicoside idrolasi (GH), che degradano il legame glicosidico tra due o più carboidrati, e gli enzimi lipolitici (esterasi e lipasi), che degradano esteri a catena alifatica, che hanno aumentato il loro impatto a livello industriale e in diverse applicazioni biotecnologiche.
In questo scenario, l'elaborato di tesi descrive nei capitoli 2 e 3 l'identificazione e la caratterizzazione di enzimi idrolitici mediante approcci metagenomici e di *genome mining*. Mentre, studi più avanzati su potenziali applicazioni biotecnologiche di GH, sono descritti nei capitoli 4 e 5. Nel capitolo 2 è stato utilizzato un approccio di tipo metagenomico per esplorare la componente microbica che popola l'ambiente estremo della solfatara Pisciarelli, con lo scopo di identificare nuove sequenze geniche codificanti per putativi enzimi attivi sui carboidrati (CAZymes). Questi enzimi sono classificati all'interno della banca dati CAZy disponibile *on line* in cinque gruppi principali: glicosil transferasi (GT), GH, esterasi dei carboidrati (CE), polisaccaridi liasi (PL) ed attività ausiliari (AA) ed ogni gruppo è diviso in famiglie e sottofamiglie sulla base di omologia di sequenza, meccanismo catalitico e similarità di struttura tridimensionale. L'ambiente solfatarico analizzato era caratterizzato da due pozze fangaie ad alta temperatura e pH acido, denominate Pool1 (86 °C, pH 5.5) e Pool2 (95 °C, pH 1.5), da cui sono stati ottenuti due *dataset* metagenomici. Dallo studio è emerso che l'ambiente era interamente popolato da microrganismi ipertermofili appartenenti al dominio degli Archaea e sono state identificate 278 sequenze codificanti per putativi CAZymes da Pool 1 e 308 da Pool2. Dai dati metagenomici di Pool2 sono state selezionate due sequenze codificanti per putative GH, che sono state espresse eterologamente e gli enzimi sono stati caratterizzati biochimicamente in dettaglio: una β-mannanasi/β-1,3-glucanasi di famiglia GH5 (sottofamiglia 19) ed una β-*N*-acetylglucosaminidasi/β-glucosidasi appartenente alla famiglia GH109. Più in dettaglio, in questo lavoro mi

sono occupato della produzione e la caratterizzazione dell'enzima appartenente alla famiglia GH109 (GH109_Pool2). Questa famiglia enzimatica era poco caratterizzata, infatti contava solo 6 membri caratterizzati. Gli enzimi appartenenti a questa famiglia presentano un meccanismo NAD^+-dipendente ed hanno attività sia sugli anomeri *beta* che *alfa* del legame glicosidico, caratteristica non canonica per le GH. Inoltre, tutti gli enzimi caratterizzati presentano un'attività *N*-acetilgalactosaminidasica. GH109_Pool2 rappresenta il primo enzima della famiglia caratterizzato da Archaea, con un'attività ottimale a 85°C, ed il primo che ha una specificità di substrato diversa rispetto a quelle riportate in precedenza, espandendo, quindi, il repertorio di attività della famiglia. I risultati ottenuti hanno evidenziato la potenza dell'approccio metagenomico nell'esplorazione di consorzi microbici che popolano ambienti estremi e nell'identificazione di nuove sequenze geniche codificanti per potenziali biocatalizzatori. Inoltre, lo studio ha sottolineato l'importanza della caratterizzazione enzimatica, unico strumento in grado di darci informazioni dettagliate e non prevedibili dalla sola analisi delle sequenze geniche.
Nel capitolo 3 è stata riportata la caratterizzazione di una nuova esterasi, identificata attraverso un approccio di tipo *genome mining* all'interno del genoma del batterio termofilo *Geobacillus thermodenitrificans.* Le analisi *in silico* hanno rilevato che l'enzima apparteneva al gruppo delle enterochelin/enterobactin esterasi, un gruppo enzimatico poco caratterizzato e conosciuto principalmente nei microrganismi siderofori patogeni. Infatti, questi enzimi fanno parte di una via metabolica utilizzata dai microrganismi siderofori per catturare il ferro ambientale. Nei microrganismi patogeni, questo meccanismo, concorre nella patogenicità, permettendogli di competere con l'ospite per l'acquisizione del ferro. La caratterizzazione della nuova esterasi (EstGtA3) ha evidenziato che l'enzima mostrava caratteristiche simili alle canoniche esterasi, con un'attività ottimale in un ampio intervallo di pH (7 – 8.5) ed alla temperatura di 60°C. Inoltre l'attività di EstGtA3 aumentava in presenza del solvente n-esano ed altri agenti denaturanti (DTT, Triton-X100, β-mercaptoetanolo), rendendola particolarmente interessante per la risoluzione di miscele racemiche e la sintesi di composti puri otticamente attivi. Questo studio ha dimostrato che anche enzimi che fanno parte di peculiari vie metaboliche, mai considerati per applicazioni di tipo biotecnologico, possono avere caratteristiche interessanti e potenzialmente sfruttabili per questi scopi. Inoltre, considerando l'enorme numero di sequenze enzimatiche depositate e non esplorate, la strategia di *genome mining* può essere considerata una valida alternativa per la ricerca di nuovi biocatalizzatori.

Nel contesto della biotrosformazione delle biomasse lignocellulosiche per la produzione di bioetanolo e prodotti ad alto valore aggiunto, nel capitolo 4 è riportata la caratterizzazione del meccanismo di reazione di tre GH nell'idrolisi di oligosaccaridi dell'emicellulosa xyloglucano (XGO) da due diverse fonti, tamarindo e vinaccia di mela. I tre enzimi, dall'archaeon ipertermoacidofilo *Saccharolobus solfataricus* ed appartenenti alle famiglie GH1 (β-glucosidasi/ β-galactosidasi), GH29 (α-fucosidasi) e GH31 (α-xilosidasi), erano stati caratterizati biochimicamente in precedenza. Tuttavia, la loro azione enzimatica e, soprattutto la loro azione combinata, non erano mai state valutate in dettaglio su substrati naturali. Nel lavoro i tre enzimi hanno mostrato la capacità di idrolizzare i substrati selezionati con alta stabilità operativa a 65 °C e pH 5.5 ed, inoltre, è stato possibile definire il loro meccanismo di reazione sugli XGO. In particolare la GH29 ha idrolizzato tutti i residui di fucosio, mentre la GH1 e la GH31 hanno mostrato una forte sinergia nell'idrolisi dei substrati. Lo studio ha evidenziato l'importanza del testare l'attività degli enzimi su substrati naturali, o su substrati di interesse industriale, in modo da valutare con attenzione il loro potenziale come biocatalizzatori.
Infine, nell'ultimo capitolo della tesi, in collaborazione con l'azienda danese Novozyme A/S, sono state caratterizzate tre nuove GH della famiglia GH109 e la loro attività è stata testata per valutarne le potenzialità nell'applicazione biotecnologica della trasformazione dei globuli rossi del gruppo sanguigno A in gruppo 0, donatore universale. I determinanti antigenici dei globuli rossi che caratterizzano il gruppo sanguigno nel sistema AB0 sono rappresentati da oligosaccaridi sulla superfice degli eritrociti, con diversa composizione per i diversi gruppi. Il gruppo sanguigno A presenta come determinante antigenico lo zucchero *N*-acetylgalatosammina, legato in *alpha* ad una struttura oligosaccaridica più complessa presente su glicoproteine e glicolipidi di membrana. Se questo zucchero viene rimosso dalla superficie di tutti i globuli rossi di un campione ematico di gruppo A, il campione di sangue viene trasformato in gruppo sanguigno 0. Gli enzimi caratterizzati (NAg68, NAg69 e NAg71) hanno mostrato una spiccata attività α-*N*-acetylgalatosaminidasica e la capacità di trasformare campioni di emazie del gruppo A in gruppo 0. In particolare, la trasformazione delle emazie operata utilizzando NAg71 è avvenuta utilizzando concentrazioni enzimatiche più basse rispetto quelle utilizzate dai migliori enzimi attualmente disponibili per questa biotecnologia. Gli studi eseguiti potrebbero rappresentare un importante punto di inizio per migliorare e permettere lo sviluppo su larga scala di questa

bioconversione, per una maggiore disponibilità di sangue universale nei centri trasfusionali e nelle banche del sangue.

SUMMARY

The development of sustainable processes and the society needs to shift towards a circular bioeconomy have increased the demand for biocatalysts in biotechnology to support industrial processes. The discovery of novel enzymes with higher activity and stability than those of catalysts already available paves the way for improving current industrial bioprocesses and developing novel applications. Microbial enzymes represent the bulk of the enzymes market, being more active and stable, producing high yield, and microorganisms represent an easily exploitable source by modern metagenomics technique. Among the more sought-after enzymes for industrial and biotechnological purposes, glycoside hydrolases (GHs) and lipolytic enzymes have increased their impact in this field. The present thesis describes different strategies to identify novel enzymes and the evaluation of several GH activities for their potential exploitation in biotechnological applications. Chapter 2 reports the exploration of microbial consortia populating two mud pools in Pisciarelli solfatara using shot-gun metagenomic approach, that led to the identification of 586 putative CAZymes. In this work I focused on the characterization of a GH109 with a previously unreported β-*N*-acetylglucosaminide/β-glucoside specificity. By using a genome mining approach, chapter 3 describes the identification of a novel thermophilic esterase (EstGtA3) from the thermophilic bacterium *G. thermodenitrificans*. The characterization revealed that the enzyme is active at 60°C and in a wide range of pH. Moreover, EstGtA3 showed an activating effect in *n*-hexane and other denaturing agents, making this enzyme suitable for biotechnological applications. More advanced studies on possible industrial applications of some GHs are reported in chapters 4 and 5. In chapter 4 the characterization of the mechanism of action of three thermostable GHs (LacS, XylS and SsαFuc) from the hyperthermophilic archaeon *S. solfataricus* on xyloglucan oligosaccharides shows the excellent operational stability at 65°C and pH 5.5 of the three enzymes. SsαFuc was able to remove all fucose residues, while LacS and XylS showed a strong synergy for the hydrolysis of these substrates. The last chapter, in collaboration with Novozymes A/S company, describes the characterization of three novel GH109 (NAg68, NAg69 and NAg71). Interestingly, these enzymes showed the ability to remove the immunogenic determinant N-acetylgalactosamine from the erythrocyte surface of group A blood, converting it into the universal donor group 0. In particular, NAg71 requires less enzyme concentration for conversion than the already available enzymes used for this biotechnological application.

ABBREVATION

- **AA**: Auxiliary activities
- **CAGR**: Compound annual growth rate
- **CDD**: Conserved domain databases
- **CE**: Carbohydrate esterases
- **CEc**: Circular economy
- **COG**: Cluster of orthologous groups
- **eDNA**: Environmental DNA
- **FACS**: Fluorescence-activated cell sorting
- **FPLC**: Fast protein liquid chromatography
- **Fuc**: Fucose
- **Gal**: Galactose
- **GalNAc**: *N*-Acetyl-galactosamine
- **GH**: Glycoside hydrolase
- **Glc**: Glucose
- **GlcNAc**: *N*-Acetyl-glucosamine
- **GT**: Glycosyltransferase
- **HPAEC-PAD**: High performance anion exchange chromatography with pulsed amperometric detector
- **IPTG**: Isopropil-β-D-1-thiogalattopiranoside
- **KEGG**: Kyoto encyclopaedia of genes and genomes database
- **LEs**: Lipolytic enzymes
- **Man**: Mannose
- **MEC**: Minimum enzymes concentration
- **NCBI**: National center for biotechnology information
- **NGS**: Next generation sequencing
- **OD_{600}**: Optical density at 600 nm
- **OTU**: Operational taxonomic units
- **PL**: Polysaccharide lyases
- **RBC**: Red blood cell
- **trRBCs**: Enzymatic treated red blood cells
- **USD**: United states dollar
- **XG**: Xyloglucan
- **XGO**: Xyloglucan oligosaccharide
- **Xyl**: Xylose
- **4NP**: 4-nitrophenol

CHAPTER 1

INTRODUCTION

1. General Introduction

Enzymes are biocatalysts that accelerate the biochemical reactions underlying the metabolism of all living beings. The word enzyme derived from the Greek *en* (meaning "within") and *zume* (meaning "yeast") and was first used by the German physiologist Wilhelm Kühne in 1878, during his studies on the ability of yeasts to produce ethanol from sugars. In the following years until to come, significant advances were made in the identification, characterization, and commercial exploitation of many enzymes [1]. Enzymes catalyze chemical reactions with excellent specificity and processivity, and over the years, their use in the industrial field has overgrown, allowing to increase the yields and quality of products, lowering the energy usage and the environmental impact. The growing development of sustainable processes has increased the demand for new biocatalysts [2]. According to Grand View Research (www.grandviewresearch.com), the global enzymatic market for biocatalysis is continuously growing. In 2019 it was estimated at 10 billion USD and was expected to grow at a compound annual growth rate (CAGR) of 7.1% from 2020 to 2027 [Enzymes Market Size & Share Industry Report, 2020-2027]. Most industrial enzymes are hydrolases, such as carbohydrate hydrolases, proteases, and lipolytic enzymes, catalyzing natural polymeric substrates' breakdown. Pharmaceutical, food, beverage, animal feed, biofuel, detergent, textile, and laundry industries benefit from the use of enzymes, and several other sectors such as natural gas conversion and fine chemicals production are recently considering their usage. Proteases dominate the global enzymes market, along with glycoside hydrolases (GH), which is the fastest-growing segment of this market in terms of CAGR. Various industrial processes use lipolytic enzymes, and their importance at the industrial level multiplies, particularly in pharmaceutical production and food industries, fields that dominate the enzymes global market [2-4].

Despite their growing importance in bioconversions, the introduction of enzymes into the industrial field is proceeding slowly. The main reasons are the limited commercial availability of some enzyme classes and the poor performance of many enzymes in the typical harsh conditions of industrial processes. In these processes, the concentrations of substrates and products are much higher than those found in nature, and the enzymes often operate on very different molecules than their natural substrates. Furthermore, the substrates and products are often hydrophobic, and organic solvents improve their solubility [4,5]. Therefore, discovering new enzymes with higher activity or better regio- and enantioselectivity than those of the already available catalysts

become indispensable. Some applications may require completely novel activities. Moreover, properties such as resistance to organic solvents, thermostability, and process robustness are needed [6]. The discovery of industrial enzymes is mainly addressed to microbial sources. Indeed, the bulk of commercial enzymes come from microorganisms due to their accessible laboratory culturing, natural abundance, and rich diversity. Compared to animal and plant enzymes, microbial enzymes have several advantages: they are more active and stable, produced high yield, and microorganisms represent an easily exploitable source by modern metagenomics technique. Metagenomics approaches allow access to whole microbiomes of virtually any environment, including extreme ecological niches, offering the access to a vast reservoir of unexploited enzymes possibly with unique structure-function properties exploitable for industrial purposes [6,7]. Among the most sought-after hydrolases the lipolytic enzymes (esterase and lipase) and glycoside hydrolases (GH), which catalyse, respectively, the hydrolysis of acyl chain ester and glycosidic bond, have increased their importance at the industrial level in several sectors [8]. The discovery, characterization and commercial exploitation of novel enzymes could be a key to the improvement and development of new biotechnology.

1.1. Discovery of novel enzymes

Identify and produce novel enzymes is the most viable alternative to access biocatalysts with new properties. It is possible to design *de-novo* enzymes and engineer already characterized enzymes to obtain biocatalysts not existing in nature, having new properties and exploitable for biotechnological applications. However, their introduction into the industrial sectors is minimal due to the slowness and high costs of the processes involved.

Nature can be considered the pioneer of protein engineering, as life in ecosystems is constantly changing, magnifying biodiversity. Nowadays, microbial diversity, for the most part unexplored, represents the prominent source to discover novel potential biocatalysts for developing new biotechnologies [9-12]. The emergence of next-generation sequencing (NGS) has allowed identifying an extraordinary amount of sequences from genomic and metagenomic datasets, exponentially expanding the availability of new biocatalysts-coding sequences. The NGS revolution began with the introduction in 2005 of pyrosequencing technology by 454-Life Science (Branford, CT, USA) that generates millions of short sequencings, called reads, in a single machine run [13]. Nowadays, different NGS technologies are developed, and mainly

classified on the basis of the length of reads generated, that can be short (50-400 bp) or long (1-100 kb) according to the second and third NGS generation, respectively [14] (Table 1).

Table 1: Example of NGS technology of second and third generation

Technology	Generation	Strength	Weakness	Reads Output
Illumina	2nd	Accuracy	High costs	50-300 bp
Ion torrent	2nd	Speed	Error rate	up to 400
PacBioRSII	3rd	Read length	High error rate	Up to 60 Kb
MinIon	3rd	Read length	High error rate	Up to 100 Kb

Despite the clear advantage of third generation NGS generating long reads, they still have a higher error rate. Indeed, the Illumina technology, despite the high costs, dominates the market of NGS thanks to its accuracy [15,16]. NGS, and the following development of metagenomics technologies, have made possible the study of environmental DNA (eDNA) from microbial consortia populating different places such as soil, seawater, organisms' gut, and extreme environments [17]. Over the past decade, enzymes identified from sequenced eDNA (metagenomic DNA) have been multiplied. Table 2 shows some of the identified and characterized GHs and lipolytic enzymes that have interesting features in different industrial sectors [22-33]

Table 2: Some GHs and lipolytic enzymes identified in the last decades exploitable for industrial application.

Enzymes	Source	Application	Ref
Endoglucanase	Soil	Detergent, Biofuels, Paper and Pulp	[18]
Endoglucanase	Hot Spring	Detergent, Biofuels, Paper and Pulp	[19]
Endoglucanase	Ovine rumen soil	Detergent, Biofuels, Paper and Pulp	[20]
Endoglucanase	Algae	Detergent, Biofuels, Paper and Pulp	[21]
Endoglucanase	Insect gut	Detergent, Biofuels, Paper and Pulp	[22]
β-Glucosidase	Hydrothermal hot spring	Detergent, Biofuels, Paper and Pulp	[23]
β-Glucosidase	Bovine rumen	Detergent, Biofuels, Paper and Pulp	[24]

β-Xylanase	Compost	Detergent, Biofuels, Paper and Pulp	[25]
α-Xylosidase	Soil	Biofuels, Food and Feed processing	[26]
α-Fucosidase	Soil	Bioethanol production	[27]
β-Galactosidase	Hot spring water	Bioethanol production and food processing	[28]
Esterase	Marine mud	Pharmaceutical, food, detergents, laundry, agro-industrial, biofuels	[29]
Esterase	Hot spring mud	Pharmaceutical, food, detergents, laundry, agro-industrial, biofuels,	[30]
Esterase	Permafrost	Pharmaceutical, food, detergents, laundry, agro-industrial, biodiesel	[31]
Lipase	Human oral microbiome	Pharmaceutical, food, detergents, laundry, biodiesel	[32]
Lipase	Fed batch reactor	Pharmaceutical, food, detergents, laundry, biodiesel	[33]

The source of the metagenomic DNA increases the possibility of finding new biocatalysts able to operate in the harsh industrial conditions. In this regard, extremophiles are source of the enzymes particularly suitable for this purpose thanks to their natural resistance to the extreme conditions. Indeed, extremophiles find their niches in environments and ecosystems that from anthropocentric perspective might not be classified as normal habitable but as “extreme”, in which there are present conditions mostly lethal for the non-extreme organisms [34]. However, the massive identification of sequences has generated a substantial discrepancy between functionally/structurally characterized enzymes and annotated sequences encoding for enzymes. The number of characterized enzymes is much lower than the total number of enzyme sequences annotated, limiting the exploitation of exciting new activities. Efforts to characterize novel enzymes are needed [35]. In order to make more accessible the efforts in enzymes characterization, the enormous amount of sequences can be analysed by the computational methods to predict the function of the encoded proteins [36]. In the following paragraphs some of these aspects will be described more into detail.

1.1.1. Metagenomics

Metagenomic is a culture-independent technique that potentially allows exploring any kind of environmental sample. The analysis of ribosomal

16S sequences from metagenomic DNA (mDNA) allows a taxonomic analysis of the entire microbial community. The 16S sequences are generally clustered into Operational Taxonomic Units (OTUs) and compared with reference databases for taxonomic analysis. By targeting the short reads to the variable regions of 16S genes or sequencing the entire 16S gene (~ 1500 bp) it is possible to achieve a taxonomic classification at the resolution of genus, species, and strains level. [44,45]. Having a picture of the microbial consortia populating the environment of interest can facilitate the selection of appropriate heterologous expression systems to use for recombinant expression of genes of interest [46,47]. Metagenomics is used for enzyme discovery by following two different strategies: sequence-based and function-based screenings. The selection of sampling location is crucial for the aim of the study, as environmental conditions (i.e., temperature, pH) will determine the characteristic of the enzymes sought [37,38].

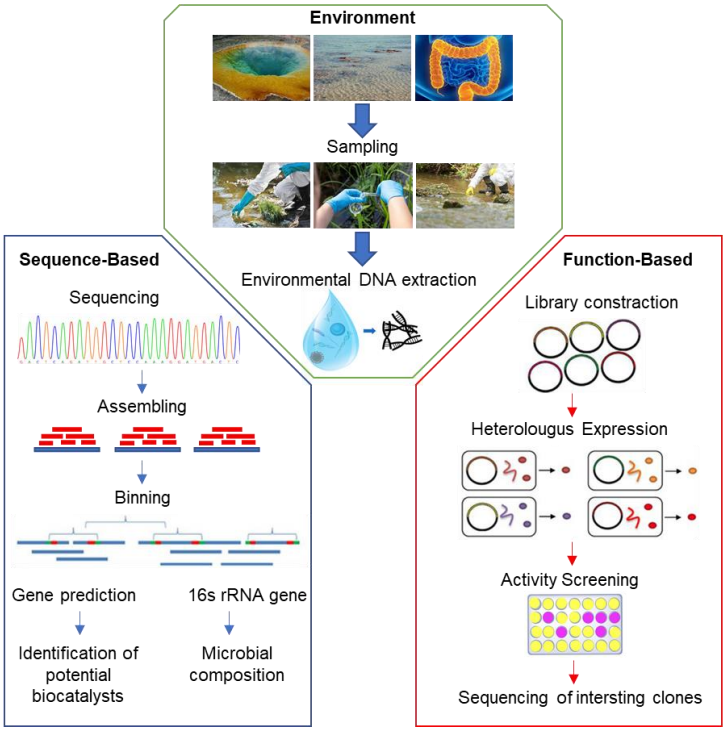


Figure 1: Framework of metagenomics approaches

In the sequence-based metagenomic approach, or *shotgun* metagenomic, consists of the directly sequencing in-depth of eDNA and bioinformatic plays a crucial role in analyzing the data [39]. The assembling process, in Fig 1, uses bioinformatics tools (i.e., MEGAHIT [40]) to join the short metagenomics *reads*, deriving from NGS, in longer

sequences called *contigs*. The assembling is followed by the *binning* process. The *contigs* are compared against database of genomic sequences and clustered into individual groups taxonomically assigned to specific species, subspecies, or genera, by using software such as MyCC, MaxBin and Autometa [41-43].

The annotation consists of the predictions of the coding DNA sequences followed by the functional annotation through homology-based screening (Fig. 1). Thus, the sequence-based metagenomic approach mainly identifies variants of already known enzyme classes, whose activities are just inferred. Only after the complete characterization of the recombinant enzymes is it possible to unequivocally assign the function. However, this approach led to the identification of novel enzymes with higher specific activity than those already available. For instance, the metagenomic analysis of Sargasso Sea and of the cow rumen allowed the identification of 70000 and 27000 novel genes for putative enzymes, respectively [48-50].

The function-based metagenomics, on the other hand, is the approach of choice to have higher chances of discovering potentially novel classes of enzymes with no homologies to known sequences. In this approach, libraries are produced by cloning the mDNA into expression vectors and that are propagated in appropriate hosts. The following step consists of activity screenings of the transformants on selected substrates (Fig. 1) [51]. The success to find novel activities is dependent on a faithful heterologous expression system. Several factors can hamper foreign DNA expression, namely different codon usage, transcriptional and translational regulatory sequences, and protein folding. More, functional-based metagenomic could be limited by the availability of high-throughput screening methods for particular enzyme classes and activities [44,52]. The majority of metagenome-sought biocatalysts were identified by this approach [54]. Ferrer et al. reported that on ~ 6100 total targets identified by metagenomic approach in the last two decades, 5800 were identified by functional screening. More, Berini et al. reported that, between 2014 and 2017, on 332 novel enzymes characterized by metagenomic studies, 273 were identified by functional metagenomics [53,54]. Moreover, to overcome some limits of this procedure, like the low hits of positive clone, novel highthroughput screenings are being developed. For example, the microfluidic-driven approach could screen over 10^7 variant per day. However, good expression systems to optimize results [55].

The functional- and sequence-based metagenomic approaches should not be considered two distinct approaches, but they can work in parallel, to analyse microbial biodiversity of the environments, the genes

encoding for interesting enzymes and their function. Metagenomics revealed several novel enzymes that are beneficial for industries. Table 3 summarizes GHs and lipolytic enzymes discovered by metagenomic approaches and patented in the last 10 years [37].

Table 3: Examples of patented GHs and lipolytic enzymes in the last decade

Enzyme	Source	Patent N°	Suggested application	Year
β-Galactosidase	Termite gut	EP2530148 A1	Food processing (lactose depletion)	2011
Cellobiohydrolase	Hot spring	EP2980212 A1	Biofuel production from lignocellulosic biomasses	2013
Esterase	Brine pool	US20160053239 A1	industrial mass-production, biopharmaceuticals, and biodiesel	2013
Lipase	Soil	CN103834626 A	industrial mass-production, biopharmaceuticals, and biodiesel	2013
Xylanase	Hot spring	EP2990482 A1	Biofuel production from lignocellulosic biomasses	2014
Endoglucanase	Hot spring soil	JP6552098B2	Biofuel production from lignocellulosic biomasses	2016
Cellulase	Soil	CN108463551A	food and feed industry, detergent, weaving, and biofuel industry	2016
Chitinase	Wetland	CN107475273B	food, medicine, agriculture, and cosmetics industries	2017
β-Glucosidase	Soil	CN107828806A	pharmaceuticals, food industries, bioethanol production, and medicine.	2017
Lipase	Soil	KR102026836B1	industrial mass-production, biopharmaceuticals, and biodiesel	2018

1.1.2. Genome mining approach

Despite the significant number of sequences available, enzyme characterization remains the essential step for validating new biocatalysts for biotechnology applications and industrial processes [56,57]. Unfortunately, current methods for enzyme characterization are by far less mechanized, and more time-consuming and cost-effective if compared to NGS and bioinformatics, thereby, lagging behind the continuous increment of genes obtained from “Big data” [57]. Novel techniques based on (ultra)high-throughput methods for protein production and purification, and enzyme activity screenings have been developed [58-62]. Nevertheless, even these methods cannot bridge the gap between annotated sequences and characterized enzymes. The enormous amount of enzyme sequences with unknown functions can be filtered and rationalized through *in silico* analysis for protein function prediction and to preselect attractive targets to support experimental testing by *genome mining*. For enzyme discovery, this

method is entirely dependent on computer technology and consists of a series of bioinformatic strategies to find coding sequences for enzyme not yet explored in genome and metagenome datasets, available in public databases [63,64]. Avoiding the sampling, extraction, and sequencing of eDNA, the genome mining shares with the shot-gun metagenomic the same *in silico* strategies for coding sequences prediction and the functional annotation, using already available data. Recently, Reichard et al., by genome mining, searched in 71 public metagenomes from 58 different geothermal hot springs for novel biomass-degrading enzymes [65]. For the functional annotation, a first-line strategy is the sequence-based method. Enzymes' function can be inferred by sequence similarity against databases of known proteins and domain databases. Sequence homology prediction by public sequence databases such as UniProt (www.uniprot.org) and NCBI (www.ncbi.nlm.nih.gov) is the broad approach used for protein function annotation. However, the prediction made by this strategy is often wrong and needs to be integrated with other criteria [66]. Matching with conserved domain databases, such as PFAM (pfam.xfam.org) and CDD (www.ncbi.nlm.nih.gov/cdd), allows to investigate the presence of highly conserved motifs or domains in the sequences of interest. This strategy considers only the conserved regions and allows to classify sequences into specific protein family and subfamily, obtaining an indication of the potential activity of enzymes [67,68]. Another approach to predict protein function is the structure-based method. The fold of enzymes is strictly correlated to its function, and proteins with similar fold often have a similar function [69]. Bioinformatic tools such as Swiss-model (swissmodel.expasy.org), CATH (www.cathdb.info), Gene3D (gene3d.biochem.ucl.ac.uk), allow to predict the structure of a protein by homology-modelling with known 3D-structures annotated in protein data bank (PDB, www.rcsb.org). In particular, the local-fold similarity, especially around the active site, could help to deduce more accurately the protein function [66,70]. Many annotated sequences have no match by homology with the on-line databases. In this regard, valuable tools have been developed for a *de novo* functional annotation by using different and complementary approaches. Sequences with no significant homology against known proteins could have motifs or domain, particular fold, or other signatures, such as the genomic context, that can be correlated to protein function. For example, sequences for Carbohydrate Active enZymes (CAZymes) use to be organized in clusters in the microbial genomes. On-line databases like KEGG (www.genome.jp/kegg/pathway.html) and the Cluster of Orthologous Group (COG, www.ncbi.nlm.nih.gov/COG) use algorithms

and statistical models to integrate all these pieces of information to functionally annotate sequences with no correlated function [66,71,72]. Attractive potential activities can also be preselected using a specific database, including sequences and biochemical features of certain groups of enzymes. Carbohydrate Active enZymes databases (CAZy database, www.cazy.org) is a data bank dedicated to annotating the enzymes modifying carbohydrates. This database accurately classifies CAZymes in families, subfamilies, and clans based on sequence homology, activity, and structure-related similarities [73]. This database allows interesting preselecting activities in genomic and metagenomic datasets, based on the similarity with classified enzymes, functionally and structurally characterized.
Bioinformatic offers a great variety of strategies to infer enzymes' function based on their sequences, structure, evolutionary history, and association with other proteins. Genome and metagenome mining constitute a powerful and complementary approach to access new biocatalysts. Moreover, the continuous characterization and classification of novel enzymes will make these methods more robust and reliable [58,66].

1.1.3. Extremophiles' biodiversity exploitation

All three domains of life include extremophilic organisms and most of them are microorganisms. The most frequent microorganisms encountered in extremophilic environments belong to the Archaea domain [74]. Extremophiles can be classified according to the required extreme condition, such as extreme of temperature, pH, high concentration of salts, radiations (Table 4).

Table 4: Extremophiles' classifications. Adapted from Fongaro et al. 2020 [74]

Extreme Condition	Growth parameteres	Extremophiles
Extreme temperature	60 – 85 °C	Thermophiles
	> 85 °C	Hyperthermophiles
	< 15 °C	Psychrophiles
Extreme pH	Low pH (< 2.0)	Acidophiles
	High pH (> 10.0)	Alkaliphiles
High salt concentration	2 – 5 M	Halophiles
High pressure	40 – 130 MPa	Piezophiles
Low water activity	aw ≤ 0.8	Xerophiles
Ionizing radiation	> 25 kGy	Radiophiles
Heavy metals	Cu, Cd, As, Zn	Metallophiles

The study of extremophiles plays a key role in different research areas, allowing to acquire knowledge on adaptation to extreme conditions, the

biogeochemical cycle of elements, and the limits of life with the implication in the developing field of astrobiology, for the study of the origins of life and the search for life in extra-terrestrial environments [75,76]. The extremophiles are considered an excellent reservoir of biodiversity, that might have dominated our planet's evolutional history [76]. The impact of extremophiles in biotechnology is widely recognized, mainly thanks to the unique features of their enzymes (extremozymes), exploitable for industrial processes. Extremozymes, catalyzing reactions under conditions where the mesophilic counterparts are completely inhibited or even denatured, find application in different field of biotechnology [77]. In the Table below examples of GHs and lipolytic enzymes from extremophiles and their applications are reported.

Table 5: Example of extremophilic enzymes and their application

Halophiles		
Enzyme	**Application**	**Ref**
α-amylase	Detergent and food industries	[78]
Xylanase	Paper and pulp industries	[79]
Cellulase	Food, laundry, textile, pulp and paper industries	[80]
Lipase and esterase	Food and nutraceutical industries, chemicals, and biodiesel production	[81]
Psychrophiles		
Enzyme	**Application**	**Ref**
Pectinase	Cheese ripening, fruit juice and wine industries	[82]
Xylanase	Dough fermentation, fruit juice and wine industries	[83]
α-amylase	Detergent, dough fermentation, desizing denim jeans, pulp bleach	[84]
β-galactosidase	Dairy industries	[85]
Acidophiles		
Enzyme	**Application**	**Ref**
α-amylase	Detergent, bread, and textile industries	[86]
Glucoamylase	Dextrose and fructose syrup, brewing of low calories beer, baking and alcohol industries	[86]
(Hyper)Thermophiles		
Enzyme	**Application**	**Ref**
Amylase	Starch hydrolysis, brewing, baking detergents, production of maltose	[87]

Cellulase	Cellulose hydrolysis for biofuels production, polymer degradation in detergents	[88]
Chitinase	Food, cosmetics, pharmaceuticals, agrochemical industries	[89]
Xylanase	Xylan degradation in pulp and paper industries and in biorefineries for biofuel production	[90]
Lipase and Esterase	Dairy, oleo chemical, detergent, pulp, pharmaceutical, cosmetic, agrochemical and leather industries	[91]

Enzymes from halophiles are active and stable in high concentration of salts, in presence of organic solvent, and at low water activity. GHs from halophiles are employed in the hydrolysis of starch and cellulose. While halophilic lipases and esterases have a wide range of applications, such as in food industries, fine chemicals, and biodiesel production [78-81]. Enzymes from psychrophiles show excellent properties in the production of fine chemicals in the detergent and food industries. A striking example is the β-galactosidase from the psychrophilic yeast *Guehomyce spullulans* that is used to produce lactose-free milk at low temperature, avoiding contamination issues [92]. Among the extremophiles, thermophiles and hyperthermophiles, the majority of which belonging to the Archaea domain, are an essential source of interesting biocatalysis (namely thermozymes) with increased stability at high temperatures, extreme pH, in the presence of organic solvents, heavy metals, and against proteolytic attack. Moreover, their exploitation in industrial processes limits the contamination risks, increases the solubility of substrates, and accelerates the rate of reactions [93]. These microorganisms thrive in several high temperature environments such as deep-sea hydrothermal vents, shallow terrestrial hot springs, and volcanic areas, in which poly-extreme conditions often coexist. As a consequence, many thermozymes are poly-extremophilic being naturally resistant to different harsh industrial conditions, such as high temperature and acid pH [94]. By starting from the classic and the most successful example of the hyperstable "Taq polymerase", nowadays, other (hyper)thermostable enzymes are produced at world-wide large-scale, such as lipolytic enzymes and GHs. Thermostable GHs has a wide range of industrial applications. Several industrial sectors perform some processes at elevated temperature and benefit from the use of thermostable GHs [95]. Commercial enzymatic cocktails for the bioethanol production, such as Cellulaclast 1.5 (Novozyme A/S), Cellic® CTec2 (Novozymes A/S) and ACCELLERASE® 1500 (DuPont) are composed mainly by cellulases and hemicellulases with optimum temperature between 50-60 °C and pH 4-6. Indeed, the

biotransformation of lignocellulose biomasses to bioethanol finds in thermostable GHs a valuable opportunity for making the bioconversion cost-competitive [96-98]. Thermostable lipolytic enzymes find application in paper, milk, leather industries, in pharmaceutical, and in biodiesel production (table 5). In food industries, thermostable esterases are used to produce fruit juices, beer, alcohol, wine, and fragrance. Moreover, these enzymes are also employed in the production of pesticides, insecticides, and nematicides in agriculture industries and in the resolution of racemic mixture [98,99]. Recently a novel esterase from a hot spring metagenomic library has been discovered showing remarkable stability and performance on the enantiomeric hydrolysis of racemic ibuprofen and naproxen [91].
Despite the clear advantages obtained by using thermostable enzymes, few of them, compared to all those identified from (hyper)thermophiles, reach industrialization. The exiting gap between producing extremozymes in lab conditions and the production at industrial level is still a drawback for their exploitation. This issue is due to difficulty in culturing hyperthermophiles on large scale. Only few of them are culturable and the laboratory equipment often are not suitable. These microorganisms usually have a low growth rate compared to the mesophilic counterpart. Thus, the overexpression of the thermozymes is performed in mesophilic hosts that guaranty fast growth rate and enough biomass for a large scale production. Nevertheless, extremophilic genes often cannot be expressed efficiently in these hosts [100,101]. For all the reasons shown above, biodiversity of (hyper)thermophilic microorganisms, and more generally of extremophiles, is for the most part yet unexplored, and their study is still in active development to improve existing bioprocesses and pave the way for novel biotechnological applications [95,96].

1.2. Carbohydrate Active enZymes (CAZymes)

With the term “CAZymes”, we refer to several classes of enzymes capable of modifying complex carbohydrates and glycoconjugates. Almost all living beings use carbohydrates for energy supplies, structural components, to mediate biological function and/or inter- intra-cellular recognition, and even for host-pathogen interaction. The complex polysaccharides represent the most abundant carbon supply on the planet. If we consider the variety of monosaccharides, the myriad of a possible linkage between them, and the modification of carbohydrates (i.e., acetylation, sulfation), glycans are the biomolecules with the broadest diversity. This diversity is reflected in the wide variety of enzymes that modify carbohydrates, and due to the

various role of carbohydrates in nature, these enzymes find applications in human health, nutrition, and biotechnology [102].
CAZymes are currently classified in classes in CAZy databases (www.cazy.org), namely GHs, glycosyltransferases (GT), carbohydrate esterases (CE), polysaccharide lyases (PL) and auxiliary activities (AA). Each class is formed by families, which are created on the basis of sequence similarities, grouping enzymes with common fold, catalytic machinery, stereospecificity, and reaction mechanism. Families may include enzymes with different substrate specificities experimentally determined. Members of the same family with high correlation between sequence and substrate specificity are grouped in subfamilies. Moreover, families with the same three-dimensional structure are grouped into clans, indicating a common origin from an ancestral gene [73]. To date (April 2021) CAZy databases ranks 170 GH, 114 GT, 41 PL, 18 CE, and 16 AA families. The growth rate of new sequences identified boost exponentially and novel families continue to be added. This classification makes easiest the identification of novel CAZymes and experimental effort in novel enzymes characterization [103].

1.2.1. Glicoside Hydrolases

Glycoside hydrolases are the most abundant enzymes in all living systems and play essential roles in treating carbohydrates [104]. The GHs grouped in the different CAZy families, can be distinguished in different way. GHs can be classified based on their action mode on the long sugar chain, in *exo*- or *endo*-glycosidases. *Exo*-glycosidases catalyze the hydrolysis of glycosidic bond from the reducing or non-reducing ends of oligo- or polysaccharides. In contrast, the *endo*-glycosidases hydrolyze the glycosidic bonds in the middle of oligo- or polysaccharides chains.

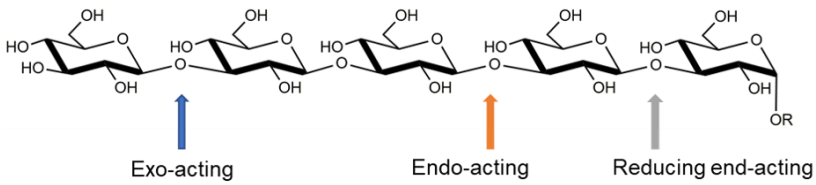


Figure 2: Example of exo-glycosidase and endo-glycosidase

Based on the stereo-specificity of the anomeric carbon involved in the glycosidic bond, they can be classified as α- or β-glycosidases (Fig. 3) or as *inverting* or *retaining*, on the basis of the reaction mechanism that may lead to different stereochemistry of the product if compared to the substrate [105].

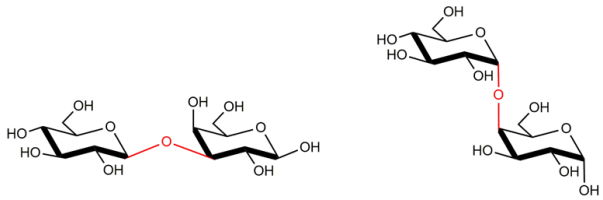


Figure 3: Example of *beta* and *alpha* glycosidic bond

Inverting GHs catalyze the hydrolysis reaction in a single-displacement mechanism, generating an inversion of the anomeric carbon in the product respect to the substrate (Fig 2). The mechanism involves two carboxylic acids, usually glutamic or aspartic acids residues: one acts as a general acid while the other acts as a general base. The general base assists the nucleophilic attack driven by water molecule, while the general acid catalyze the protonation of the aglycon, resolving the reaction with the formation of products.

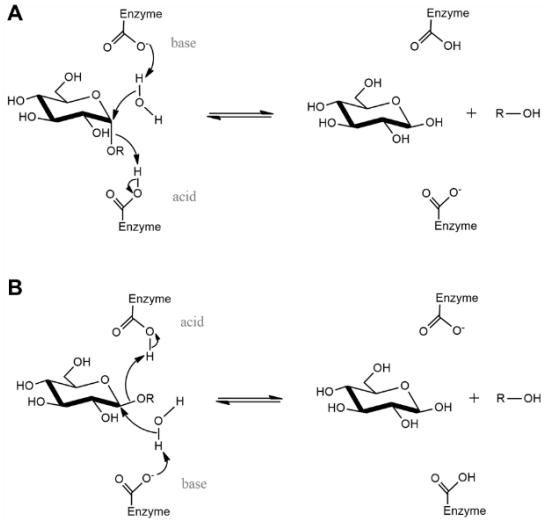


Figure 4: *Inverting* mechanism. A: *alpha* inverting B: *beta* inverting

By contrast, the reaction mechanism of *retaining* GHs proceeds via a double displacement mechanism involving a covalent glycoside-enzyme intermediate, with the retention of anomeric carbon in the products. In the first step (named *glycosylation* step), one carboxylic residue, acting as general acid, catalyzes the protonation of the aglycon (named *leaving group*), while the carboxylate nucleophile attacks the anomeric C1 atom forming the glycosidic bond. As a result, a covalent bond between the carboxylate and the anomeric carbon of the substrate is formed, leading to the *glycosyl-enzyme intermediate* and explaining

why this step of the reaction is named the *glycosylation* step. In the second step (the *de-glycosylation* step), the carboxylic group that in the previous step was acting as a general acid, now acts as general base and deprotonate a water molecule which attacks the anomeric centre of the *glycosyl-enzyme*, allowing the release of the product with the same anomeric configuration of the substrate (Fig. 5) [106].

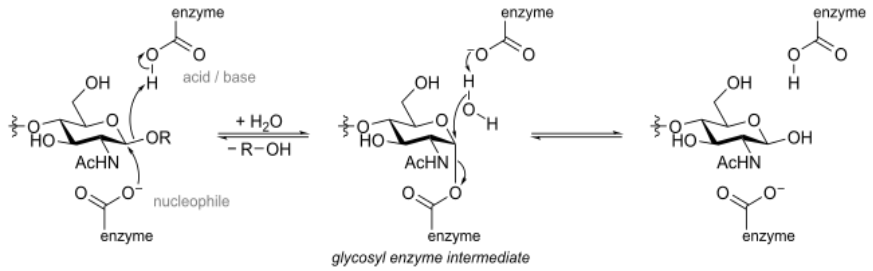


Figure 5: *Retaining* mechanism

Some families of GHs do not follow the classic mechanism described by Koshland. An example are the enzymes from GH4 and GH109 families, that have a particular reaction mechanism involved NAD^+ molecules as a molecular cofactor. The enzymes of both families can recognize as substrate both α- and β-anomer of the glycosidic bond; however, the enzymes of the two families differ firstly for overall protein fold, substrate specificity and unlike the GH4, GH109 enzymes do not require metal ions in their reaction mechanisms [107].

1.1.2.1. GH109 family

The GH109 family has been very poorly characterized so far. To date all GH109 enzymes identified, come from microbial sources and only 7 characterized members are reported in CAZy database. As mentioned above, these enzymes do not follow the classical Koshland double displacement mechanism, but, like the enzymes of family GH4, they have an unusual NAD^+ dependent reaction mechanism of hydrolysis and are unique in their ability to cleave both α- and β- linked glycosides. By multi-sequence homology alignment and molecular dynamic simulation, a histidine in a conserved and flexible G-G-H-G-G motif has been identified as the catalytic acid-base conferring to the enzymes the ability to have both α-*retaining* and β-*inverting* activities, with two proposed mechanisms [107,108] (Figure 6, mechanism of GH109A from *Akkermansia muciniphila* [108], Scheme A and B).

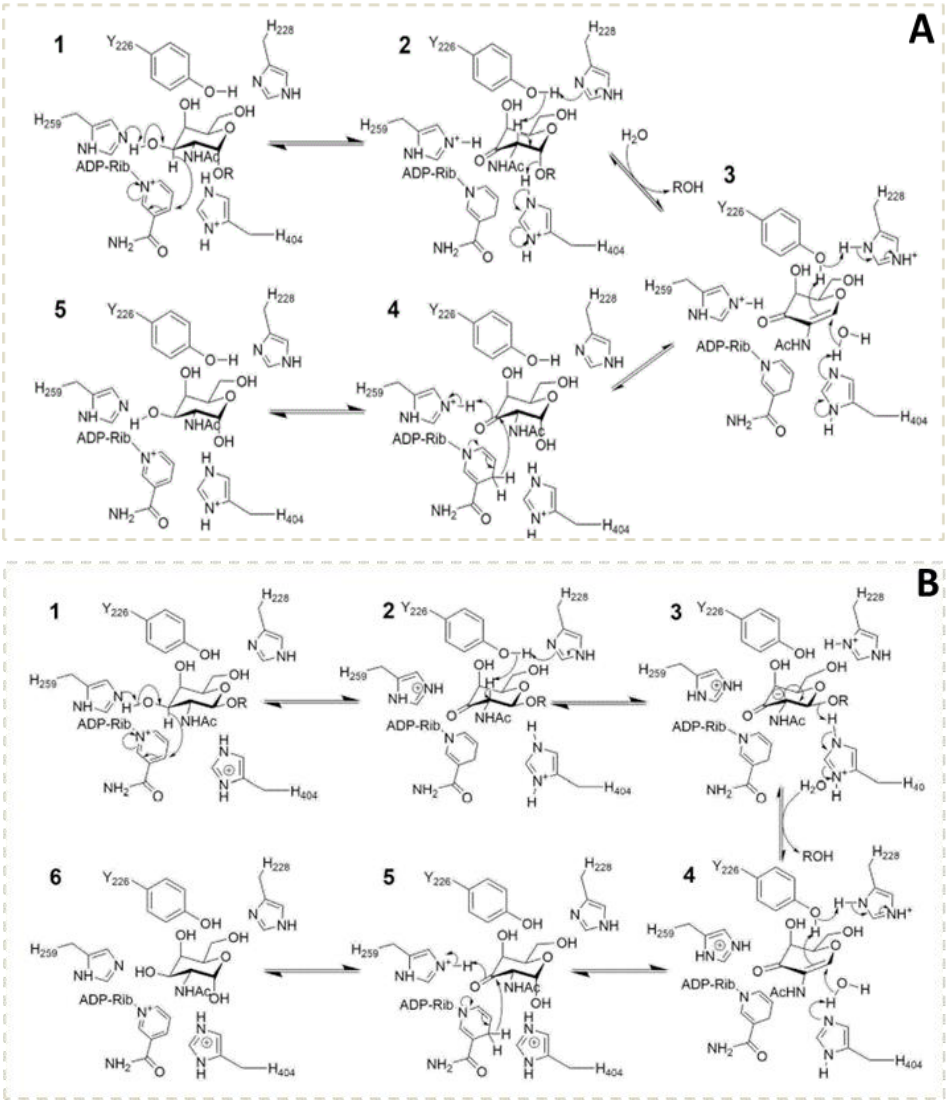


Figure 6: Proposed α-retaining (A) and β-inverting (B) mechanisms of GH109, from Teze et al. [108]

In both mechanisms, NAD^+ molecule oxidizes the C-3 of the glycon moiety of the substrate assisted by H259 base at C3-OH, oxidizing the C3-OH to a carbonyl. Deprotonation of C-2 by the catalytic base H404 with the parallel elimination of the glycosidic oxygen, generates a 1,2-unsaturated intermediate. The addition of water to the Michael-like acceptor across the C1-C2 double bond mediated by catalytic H404 (that now acts as an acid, assisted by Y226 and H228), and reduction of the resulting ketone by the NADH molecule (which returns to the initial NAD^+ state) complete the reaction (Fig. 6) [107, 108].

The NAD^+ cofactor is tightly bound, and the characterized enzymes keep it once purified in recombinant form [107,108]. The *N*-acetylgalactosaminidase activity is the only activity reported in this family in the CAZy database. From a biotechnological point of view, α-*N*-acetylgalactosaminidase activity could have potential application in the enzymatic removal of blood group AB0 antigens to develop universal red blood cells (RBCs). Indeed, the first characterized enzymes of this family (NagA) was already successful tested on A-type RBC, with the removal of the corresponding antigen from RBC. However, high amounts of enzyme and very specific reaction conditions were still needed for conversion, such as pH and low ionic strength [107]. This argument will be extensively discussed in the last chapter.

1.3. Enzymes and Circular Bio-Economy

(Poly)saccharides-degrading enzymes are employed in several sustainable bioprocesses like the production of biofuels and value-added products (i.e., chemicals building blocks) from renewable feedstock. The exploitation of natural biomasses is one of the most promising strategies to overcome the dependence from fossil resources, which have increased the pollution and distorted the natural balance of the so named greenhouse gasses. Photosynthesis converts atmospheric carbon dioxide (CO_2) into plant biomass, which, over a million year process, becomes oil, coal, and natural gas into geological reservoirs. The energy dependence on these fossil fuels returns CO_2 to the atmosphere, but in five orders of magnitude shorter time if compared to fossilization process. Therefore, the petrochemical carbon cycle results in the depletion of natural resources and an unbalanced process. The growing concern for climate change, the dwindling of natural resources, and massive environmental pollution motivate the increasing need for a shift towards a Circular Economy (CEc) [5]. CEc aims to the sustainable use of the natural resources and the recycling and reuse of waste products. Moreover, bio-based economy (bio-economy) aims to use renewable biomasses to drastically reduce fossil resource dependence towards a balanced carbon cycle safer for the environment [109]. Circular economy and bio-economy can be considered two sides of the same coin. The circular economy attempts to reshape the linear model of "take-make-use-dispose" into a circular model in which resources can be reused or biodegraded. This minimizes harmful waste and encourages reuse and restoration. On the other hand, the bio-economy takes care of the production and the use of biological processes for the supply of goods and services in a sustainable way, allowing the restoration of fossil resources. Merging

these two concepts has led to the term "circular bio-economy" for a more sustainable economy that needs to be circular [110]. Our planet needs to adopt this strategy and to achieve the “Carbon Neutrality” by balancing the CO_2 emission with its removal. The climate neutrality represents the point at which the emission of greenhouse gases does not exceed the capacity of the Earth to absorb such emission [111]. Many countries around the world have stipulated deals to achieve this goal. Europe, with the “European Green Deal”, has established that will be done the zero emission by 2050 [112].
These premises led to facilities, named biorefineries, integrating processes for the conversion of natural biomass to fuels, power, heat, and value-added chemicals, which identified in lignocellulosic biomasses one of the low-cost and uniquely sustainable resources offering the production of numerous bio-products alternative to those of the oil industry. The conversion of plant biomasses in biofuel and value-added products could be the basis of a carbon-neutral bio-based economy [5]. In this context, GHs play a key role in the field of renewable energy as they are employed in first- and second-generation biorefineries in the saccharification step of pretreated lignocellulosic biomasses to produce monosaccharides to be fermented in biofuel and plastic precursors [113].

1.3.1. GHs and Biorefineries

First- generation biorefineries use corn or sugar cane as feedstock to produce biofuels. However, concerns exist about the impact on biodiversity, land use, and competition with food crops. Therefore, first generation biofuels appear unsustainable because of the potential stress that their production places on food commodities. Otherwise, second-generation biorefineries use lignocellulosic biomasses. These feedstocks, especially those growing on marginal lands, are low-cost (i.e., primary component of agriculture wastes), not in competition with crops for food, and represent the most abundant renewable resource of energy on Earth. Nevertheless, their complex composite structure, consisting in lignin, cellulose, and hemicellulose, makes them recalcitrant to hydrolysis [114]. Cellulose is the most abundant polymer in lignocellulose formed by β-D-glucopyranose units linked via β -(1,4) glycosidic bonds, that are arranged together to form microfibrils. Hemicelluloses are a heterogeneous group of polymers, strictly linked to cellulose microfibrils, and are composed of various 5- and 6-carbon sugars, such as arabinose, galactose, glucose, mannose, and xylose. Lignin is a complex heteropolymer of phenylpropanoid building units (*p*-

coumaryl, coniferyl, and sinapyl alcohol), linked to cellulose microfibrils through the hemicellulose (Fig. 7) [115].

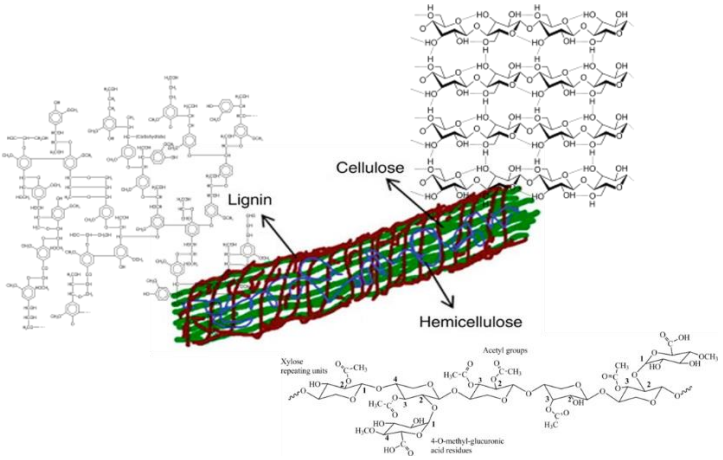


Figure 7: Representation of lignocellulose microfibril

Thus, the complete bioconversion of these feedstocks requires a complex process whose basic steps are: (1) mechanical or chemical pretreatment; (2) saccharification step by acid or enzymatic hydrolysis to break down polysaccharides into simple sugars (hexoses and pentoses); (3) microbial fermentation of the monosaccharides to ethanol or chemical building blocks; (4) separation and concentration of the final bioproducts [116].

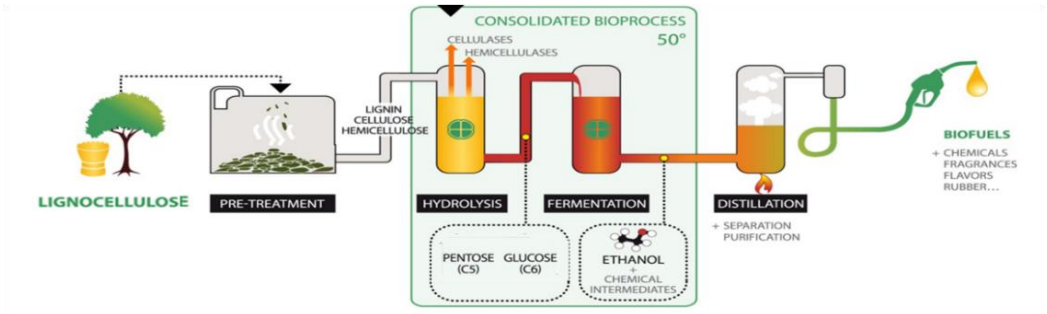


Figure 8: Schematic illustration of second-generation biorefinery process

Thereby, after the pretreatment of the row material, cellulose and hemicellulose are released from lignin, and they must be saccharified in monosaccharides (Fig 8). In an efficient lignocellulosic second-generation biorefinery the enzymatic saccharification is an essential step. Enzymatic hydrolysis has numerous advantages than acid treatment, such as specificity, low by-products formation and environmental sustainability. The main obstacles for large-scale

enzymatic saccharification include high enzyme costs, and low catalytic efficiency and operational stability of conventional (hemi)cellulolytic enzymes. These hampers can be overcome by using enzymes from (hyper)thermophiles. Thermozymes well meet the conditions that improve the bioprocess: they operate at high temperature, allowing a more solubilization of the substrates and avoiding the expensive cooling steps; moreover, they could operate in the presence of solvents and avoid the risks of contaminations [117,118].

1.4. Lipolytic Enzymes

Along with CAZymes, the lipolytic enzymes (LEs) have acquired much importance in several industrial sectors as described in the previous paragraphs. LEs, including esterases and lipases, catalyze the hydrolysis of ester bonds between alcohols and carboxylic acids in aqueous media. In addition, in organic solvents, they catalyze the synthesis of ester bonds by esterification or transesterification. LEs are ubiquitous enzymes present in all three domains of life; in particular, LEs from microbes can catalyze different reactions without the need for cofactors and are active and stable in organic solvents. These enzymes attract much attention in the industrial field, thanks to their robustness, activity in a wide range of pH, and they are also chemo- regio- and stereo-selective, finding application in the synthesis of optically pure compounds [119,120]. Lipase and esterase show a low primary sequence homology, but their tertiary structure is highly conserved. They present an α/β hydrolase fold consisting of eight β-sheet connected by six α-helix (Fig. 9). In the active site it is present the catalytic triad Ser-Asp/Glu-His, with the Ser generally embedded in the consensus motif G-X-S-X-G [121].

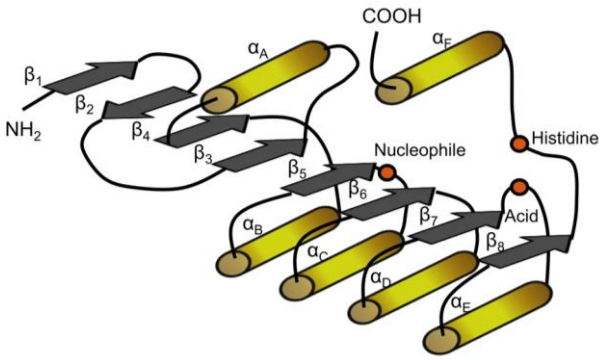


Figure 9: α/β hydrolase fold representation

Based on amino acid sequences and biological proprieties, LEs was first classified by Arpigny and Jaeger in eight families [122]. During the

years, several families were added and, recently, Hitch and Clavel described and classified LEs in 35 families and 11 subfamilies, based on novel criteria of sequence similarities and biochemical features [123].
Lipases and esterases differ in substrate specificity. Esterases act on a water-soluble ester with an acyl chain shorter than twelve carbon atoms (12 C). By contrast, Lipases are active on water-insoluble esters longer than 12C. Moreover, esterases follow classical Michaelis-Menten kinetic behaviour, while lipases generally show a sigmoidal kinetic profile due to the phenomenon of interfacial activation. At low substrate concentrations, the enzyme activity is low due to a hydrophobic lid near the active site that keeps the enzymes in a “closed” form. The increase of the substrate concentration led to the formation of emulsion in solution, which allows the opening of the lid and to the enzymes to reach their maximum activity [120,124].

1.4.1. Esterases

All microbes, such as bacteria, fungi, and archaea, produce esterases that are involved in several metabolisms of endogenous or exogenous compounds. Thermostable esterases, due to their resistance, are most required, also thanks to the ability to operate even at low temperature [125]. According to the classification of LEs, microbial esterases are classified into 13 different families and show a broad range of substrate specificity. The classification also includes esterases active on carbohydrates, namely acetyl xylan esterase (E.C 3.1.1.72) and feruloyl esterase (E.C 3.1.1.73), that are also classified in CE classes of the CAZy database [123]. The mechanism for hydrolysis or ester formation is composed of four phases (Fig. 10).

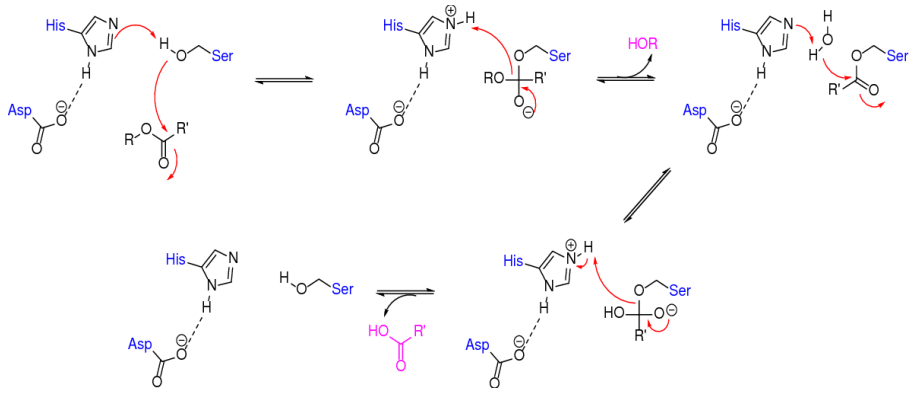


Figure 10: Esterase reaction mechanism.

The substrate is bound to the active serine, giving a tetrahedral intermediate stabilized by His and Asp catalytic residues. Subsequently, alcohol is released, and an acyl-enzymatic complex is formed. The attack of the nucleophile (water in hydrolysis, alcohol, or ester in transesterification) creates a tetrahedral intermediate and, after the resolution, produces the product (acid or ester) and free enzyme (Fig 10) [120].

1.5. Aim of work

The identification and the characterization of novel biocatalysts could represent a key to improve the exiting bioprocesses and develop novel applications. The aim of this thesis was to identify novel enzymes by two different strategies, metagenomic and genome mining approaches and to address them toward biotechnological applications. Chapter 2 reports the analysis of microbial consortia populating the extreme environment of Pisciarelli solfatara by a sequence-based metagenomic approach. The environment was characterized by two mud pools with high temperature and acid pHs. This work aimed to explore the biodiversity of the hyperthermophilic microorganism populating the pools to identify sequences for novel CAZymes exploitable for industrial purposes. In this chapter I focus my efforts on the characterization of a novel GHs of family GH109. Chapter 3 reports on the identification of a novel esterase by genome mining approach of the thermophilic bacterium *G. thermodenitrificans* in order to biochemically characterize the enzyme and explore its applicability for industrial purposes. In the frame of lignocellulose biomasses conversion, in Chapter 4 the ability of three GHs from the hyperthermophilic archaeon *S. solfataricus* to hydrolyse xyloglucan hemicellulose oligosaccharides was analysed. Finally, in Chapter 5 I analysed the ability of three GH109 enzymes to remove the immunodominant antigen *N*-acetylgalactosamine from the group A red blood cells surface in order to obtain the universal donor group 0.

CHAPTER 2

2. Cazymes identification through metagenomic of Pisciarelli solfatara

2.1. Overview

This chapter describes the production and characterization of a new archaeal GH109, identified in metagenomic datasets of Pisciarelli solfatara. This research activity, performed during my PhD thesis, is part of the wider work aimed to analyse the hyperthermophilic microbial communities populating the two pools of solfatara Pisciarelli (Pool1 and 2) and to identify the entire set of CAZymes (CAZ*ome*) in these extreme environments by a sequence-based metagenomic approach, published in Strazzulli et al. (2019) (Appendix 7.4.1). The analysis of the microbial community of the two main mud pools revealed that this solfatara is inhabited by hyperthermophilic Archaea, most of which have never been classified so far. This study had allowed the discovery of many sequences encoding for putative hyperthermophilic Cazymes. The characterization of a novel β-mannanase/β-1,3-glucanase of subfamily GH5_19, and of the first GH109 from hyperthermophiles are reported. In this framework, in the next paragraphs of this chapter, I will briefly report on the main results published in the co-authored paper, then I will describe into detail the recombinant expression and biochemical characterization of the new GH109 identified.

2.1.1. Pisciarelli Solfatara

Pisciarelli Solfatara is a hot spring in the center of the Campi Flegrei volcanic complex in the south of Italy. This volcano is a collapse caldera about 12 km in diameter, located in the north of Naples metropolitan city, and represents one of the most active volcanos in Europe and one of those located in the most populated area in the world [126]. Pisciarelli Solfatara extends over an area of 800 m^2 showing a high secondary volcanic activity (fumarole, geyser, and high-temperature geothermal water), in which two main mud pools, named Pool 1 and Pool 2 are present. Although the two pools are very close to each other's, they present different chemical-physics characteristics: Pool 1 has a temperature of 86°C and pH 5.5, while Pool 2 shows a temperature of 95°C and a pH value of 1.5 [127].
In the last 15 years the geothermal activity is continuously raising up, with an increased fumarolic discharge, formation of boiling pool and new geyser-type vents. This location represents a unique ecological niche for study the evolutionary change and adaptation of microbes at extreme conditions and a great opportunity to exploit this complex microbial population to find interesting biocatalysts [128].

2.1.2. Archaea

Archaea are microorganisms, for the most part living in extreme conditions, representing one of the three domains of life. Their discovery is due to the German biologist Carl Woese and his collaborators, who in 1977 have performed one of the most interesting microbiology discoveries in the past century, providing the first evidence that cellular life was composed of three distinct types of organisms: Archaea, Bacteria, and Eukaryota, by phylogenetic taxonomy of 16S ribosomal RNA (Fig 11) [129].

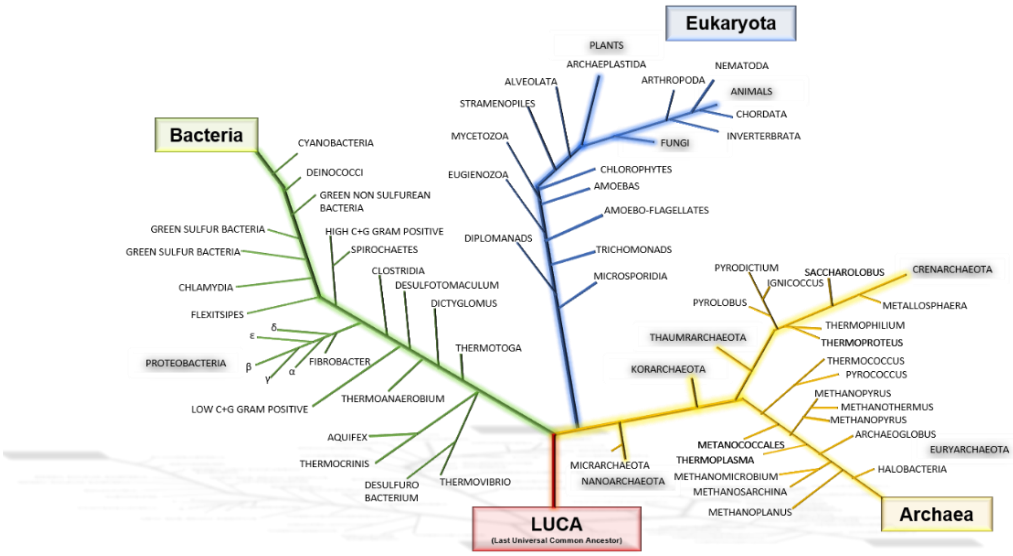


Figure 11: Graphic representation of phylogenetic three of life

Archaea share traits with both Bacteria and Eukaryota: some metabolic pathways are common to Bacteria, whereas DNA replication, transcription, and translation mechanisms share several aspects to those of Eukaryota. Moreover, they have some peculiar traits, such as the ether linkage of membrane lipids (ester linkage in Bacteria ad Eukaryota), and they are the only life forms known able to produce methane. Although some archaea live in non-extreme conditions, including the human intestine, their majority thrive in extreme environments. [130,131].

Hyperstable enzymes from Archaea have gained much attention in biotechnological applications [132]. However, Archaea are markedly less studied than Bacteria and Eukaryotes, representing an underexploited source of novel (hyper)thermozymes particularly useful for biorefineries. Indeed, few enzymes are already used at the industrial

level also due to the low yield production in heterologous systems on a large-scale [133,134].

2.1.3. Discovery of hyperstable carbohydrate-active enzymes through metagenomics of extreme environments (Results summary from Strazzulli et al. 2019. Appendix 7.4.1.)

In this study the metagenomic DNA of the microbial communities populating the two pools of solfatara Pisciarelli (Pool1 and 2) was analysed and the CAZ*ome* has been determined. Samples from Pool1, which consisted mainly of hot muddy sediments, gravel, and water, were taken from the pool's surface. Pool2 was sampled by scraping the side of the pool submerged by a clear mixture of mud/water with a spoon and consisted mainly of gravel. The eDNA was extracted from the two samples and deep sequenced with HISeq200 (Illumina) performed at Beijing Genomics Institute (BGI-Shenzhen). The community diversity was analysed by assigning reads to known microorganisms of the NCBI nucleotide database and protein databases (*NT* and *NR*, respectively) (Fig 12).

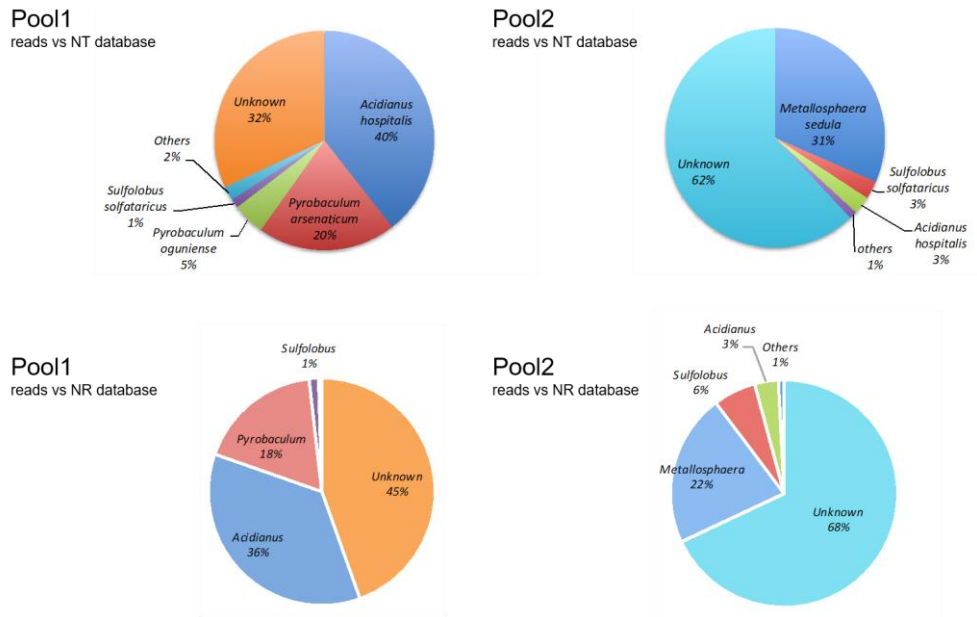


Figure 12: Read composition of the metagenome of solfatara Pisciarelli. The percentage of the reads found in Pool1 and Pool2 classified at the genus level are reported. NR, NCBI 'not-redundant' database; NT, NCBI nucleotide database.

The 32% and 45% of the reads obtained from Pool1 had no match with *NT* and *NR* databases. All the other reads belonged to Archaea and only a tiny percentage to viruses (0.17 %) and bacteria (0.11%)

(Fig. 12). From Pool2, instead, the majority of the reads obtained had no match in both *NT* (62%) and *NR* (68%). The remaining reads belonged mostly to Archaea (37%) and only a tiny part to Archaeal viruses (0.36%) (Fig. 12). These data demonstrated that the two pool samples displayed a different microbial composition, probably due to their different temperature and pH conditions. Thus, a functional analysis of the metagenomic data was performed in order to explore the metabolic biodiversity of the microbial population. From the assembling process, using MEGAHIT, 84% and 87% of the reads, from Pool 1 and 2 respectively, were successfully aligned to obtain contigs [40]. These latter were clustered with MyCC software identifying 17 and 11 clusters in Pool1 and Pool2, respectively. The open reading frames (ORFs) prediction was made by Prodigal software [135]. The number of complete ORFs identified in the whole assembly was 14934 and 17652 in Pool1 and Pool2, respectively, which were then analysed against KEGG and COG databases for the functional annotation [71,72]. The functional analyses revealed that the highest number of annotated sequences belonged to carbohydrate metabolism, representing almost 2% of the total ORFs. It has been reported that CAZymes range between 1% and 5% of all the ORFs in all taxa genomes [73]. However, no survey of CAZymes from the microbial communities populating hydrothermal environments was currently available. Nevertheless, if compared to the much more complex human gut metagenome, in which the putative ORFs that encode CAZymes corresponded to 2.6% of total ORFs, this value is remarkably high [136]. This suggested that the solfatara Pisciarelli represents a unique ecological niche for the discovery of new thermophilic CAZymes. In order to investigate the different classes of CAZymes present, the amino acid sequences identified in Pool1 and Pool2 samples were compared to the sequences annotated in CAZy database [73]. From the analyses, a total number of 278 and 308 putative CAZymes were identified in Pool1 and 2, respectively. Among them, after GTs (60 % in Pool1 and 56 % in Pool2), the second most abundant CAZymes identified were the GHs (36 % and 38%, respectively in Pool 1 and 2) (Table 6).

Table 6: CAZymes in solfatara Pisciarelli. Families with no previously characterized members are highlighted in bold

Family	Pool1	Pool2
GT2	62	65
GT4	89	81
GT5	9	11
GT20	2	0
GT21	3	5
GT35	3	5

GT66	5	6
GH1	6	5
GH2	0	1
GH3	3	2
GH4	2	0
GH5	1	4
GH12	4	8
GH13	7	10
GH15	16	27
GH29	0	1
GH31	7	9
GH36	1	5
GH38	13	8
GH57	15	12
GH65	1	0
GH77	1	0
GH78	3	1
GH99	1	0
GH109	1	1
GH116	14	20
GH122	3	3
GH130	1	0

Several GH families in Pool1 and 2 have no Archaeal characterized members (GH2, GH65, GH78, GH99, GH109 and GH130) (Table 6). Among the sequences encoding for putative GHs, to shed light on the CAZ*ome* of Pool2, two sequences were selected belonging to two different contigs of cluster 9, which from *in silico* analysis may originate from a novel microorganism not yet identified. In particular two putative enzymes from GH5 and GH109 families, were produced and characterized.

GH5 enzyme from Pool2 (GH5_Pool2) is a member of subfamily GH5_19 in which only the bacterial β-mannosidase from *Thermotoga thermarum* was characterized [137]. The coding sequence of GH5_Pool2 was amplified by PCR from Pool2 metagenomic DNA and cloned and expressed in *Escherichia coli*. The GH5_Pool2 gene encodes for a protein of 620 amino acids with a hypothetical molecular mass of about 71.3 kDa. The recombinant protein was purified through two simple steps with a final yield of 65.3% (2.2 mg L^{-1} of culture) and a purity of 95%. The molecular mass of the native GH5_Pool2 was 271 ± 13.5 kDa, as determined by size-exclusion chromatography, indicating that the enzyme is a tetramer in solution. GH5_Pool2 was optimally active on 4NP-βMan at pH 5.5 while maximum activity (0.4 U mg^{-1}) was maintained at temperatures from 65 to 85°C, indicating a broader temperature optimum. The enzyme was 100% active on 4NP-βMan for more than 22 h at 65 and 85 °C and, at 95 °C, it maintained 50% of activity for 4 h. The enzymes showed a β-mannanase and β-

1,3-glucosidase activity. In order to determine the substrate specificity, GH5_Pool2 was assayed on different aryl-glycosides, oligo-, and polysaccharides (Table 7).

Table 7: Steady-state kinetic constants of GH_Pool2 on different substrates

Substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1} s^{-1})
Mannopentaose	3.13 ± 1.20	32.90 ± 6.03	10.5
4NP-βGlc	9.90 ± 2.40	0.76 ± 0.06	0.076
4NP-βMan	0.42 ± 0.07	5.70 ± 0.45	13.75

According to these results, GH5_Pool2 should be classified as a β-mannanase and is the first thermophilic members from the subfamily GH5_19, possibly belonging to the Archaea domain.

2.2. Characterization of GH109 enzymes from Pool2

As of May 2019, the CAZy database listed 713 entries in GH109 family, exclusively among bacteria with α-N-acetylgalactosaminidase as the only activity reported. The sequences of GH109 enzyme from Pool2 (GH109_Pool2) was not annotated in the CAZy database [107, 73]. Considering the low number of characterized enzymes belonging to the GH109 family (6 members, May 2019), we decided to clone, express in a heterologous host, and characterize the putative GH109. The gene of GH109_Pool2 encodes for a protein of 364 amino acids with a calculated molecular mass of 41.26 kDa. GH109_Pool2 showed the conserved domain of the GH109 family at the *N*-terminal region (from Ile2 to Arg115). The deduced protein sequence was analyzed through Clustal Omega multialignment (www.ebi.ac.uk/Tools/msa/clustalo/) with other members of the family revealing that it was distantly clustered from the characterized α-*N*-acetylgalactosaminidase (data not showed). The protein was produced in *E. coli* with an N-terminal 6x(His) tag, but the resulting recombinant protein was insoluble. In order to obtain a soluble recombinant protein, GH109_Pool2 has been cloned with different fusion tags (Table 8) that could help the folding of fused proteins during heterologous expression (Gupta et al. 2016).

Table 8: pHTP' vectors with different fusion tags

Vector	Fusion Tag (N-terminal)	Result
pHTP1	His_6-SSGPQQGLR	Insoluble
pHTP2	Thioredoxin (Trx)-MGSS- His_6-SSGPQQGLR	Insoluble
pHTP8	Maltose binding protein (MBP)-MGSS-His_6-SSGPQQGLR	Insoluble
pHTP10	N-utilization substrate A (NusA)-MGSS- His_6-SSGPQQGLR	**Soluble**

pHTP16	*R. flavefanciens* cellulosomal protein (Rf1)-MGSS-His_6-SSGPQQGLR	Insoluble
pHTP17	*R. flavefanciens* cellulosomal protein (Rf17)-MGSS-His_6-SSGPQQGLR	Insoluble

The recombinant protein resulted soluble by using the N-terminal tag of NusA followed by 6x(His) (Table 8). The expression was performed by using *E. coli* BL21 (DE3) Star, induced at 0.5 OD_{600} with 0.5 mM IPTG for 16 hours at 20°C. The purification of the recombinant GH109_Pool2 was carried out by affinity chromatography and a subsequent size-exclusion chromatography with a final yield of 0.3 mg L^{-1} of 95% pure enzyme. The single band observed by SDS/PAGE showed a molecular mass of about 98.43 KDa corresponding to the monomer of GH109_Pool2 linked to NusA. Size exclusion chromatography analysis in the native condition revealed a molecular mass of 776.24 KDa, indicating that the enzyme is an octamer in solution (Fig. 12).

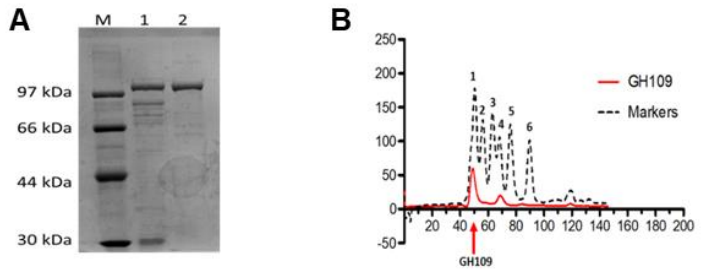


Figure 12. A: SDS-page GH109_Pool2, line 1: HisTrap elution, line 2: size exclusion chromatography. **B:** Size exclusion chromatogram, molecular weight markers, 1. Thyroglobulin 669 KDa, 2. Apoferritin 443 KDa, 3. B-Amylase, 4. Alcohol Dehydrogenase, 5. Albumin bovine serum, 6: Carbonic Anhydrase.

The preliminary characterization of GH109_Pool2 at different pHs and temperatures revealed that the enzyme has optimum activity at 85 °C and in sodium phosphate buffer at pH 8.0 (1.44 ± 0.04 U mg^{-1}) (Fig. 13A and B). The enzyme activity on different aryl-glycosides (5 mM concentration) at 75 °C and in 50 mM sodium phosphate pH 8.0 showed that it was active only in the presence of 1.5 mM NAD^+. In these conditions, the recombinant enzyme was able to hydrolyse different aryl substrate (Table 9)

Table 9: Specific activity of GH109_Pool2 on different aryl substrates

Substrate	Specific Activity
4NP-αGlc	0.33 ± 0.01 U mg^{-1}
4NP-βGlc	0.59 ± 0.02 U mg^{-1}
4NP-αGlcNAc	0.52 ± 0.01 U mg^{-1}
4NP-β-GlcNAc	0.75 ± 0.02 U mg^{-1}
4NP-βMan	0.0019 ± 0.001 U mg^{-1}
4NP-βGal	0.0018 ± 0.001 U mg^{-1}

On the other hand, no activity was detected on 4NP-*N*-acetyl-β-galactosaminide (4NP-βGalNAc), 4NP-αGal, 4NP-αMan and in particular on 4NP-αGalNAc, which is the only known substrate of GH109 enzymes.

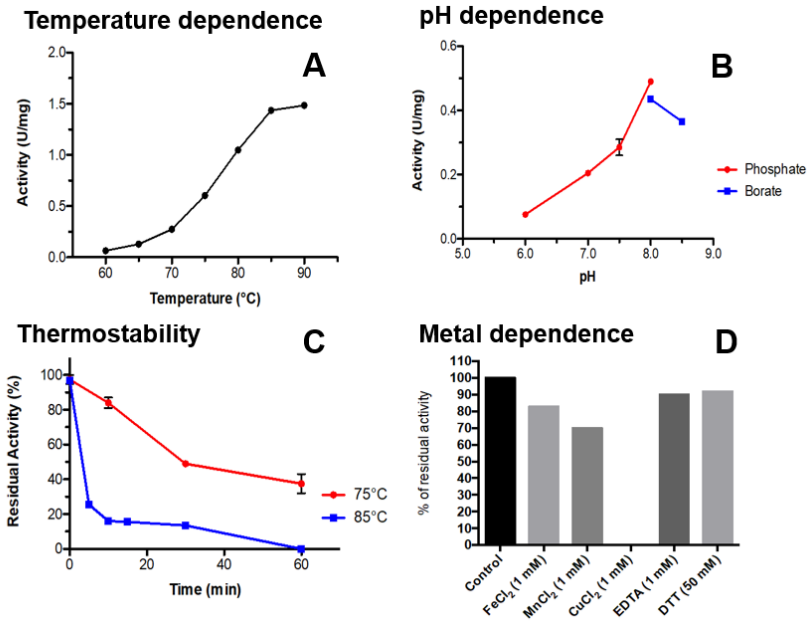


Figure 13. A*:* Temperature dependence; the assays were performed in the presence of 1.5 mM NAD^+, 5 mM 4NP-βglc, pH 8.0. **B:** pH dependence; the assays were performed in the presence of 1.5 mM NAD^+, 5 mM 4NP-βglc, at 75°. **C:** Thermal stability in standard conditions. **D:** Metal ions, metal chelating agents and DTT influences on GH109_Pool2 activity, assays were performed in standard conditions at 85 °C.

The enzyme activity at 85 °C in optimal conditions is not influenced by the presence of divalent metal ions, metal chelating agents and DTT (Fig. 13D). By contrast, $CuCl_2$ inhibited the enzyme activity, similarly to what observed for the GH109 from *Elizabecthkingia meningosepticum* [107]. To investigate the dependence from NAD^+, the enzyme was assayed on 4NP-βGlc (5 mM) at different NAD^+ concentrations, showing activation from 0.05 mM with a maximum of activity at 0.5 mM and a Kd of 32 ± 5 μM.

To investigate the substrate specificity of GH109_Pool2, the steady-state kinetic constants on 4NP-α and β-glucosides and *N*-acetyl-glucosides were determined at 75 °C. The k*cat* values were similar among all tested substrates, but the lower K_M values measured on 4NP-β-anomers (0.12 mM) made 4NP-βGlcNAc and 4NP-βGlc the preferred substrates (Table 9).

Table 10: Kinetic constants of GH109_Pool2

Substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1} s^{-1})
4NP-β-Glc	0.12 ± 0.01	4.75 ± 0.004	39.58
4NP-β-GlcNAc	0.12 ± 0.01	6.11 ± 0.01	50.91
4NP-α-Glc	1.48 ± 0.18	3.28 ± 0.02	2.22
4NP-α-GlcNAc	0.87 ± 0.06	4.98 ± 0.01	5.72

The kinetic characterization showed that this first archaeal GH109 enzymes differed significantly from the other known members of this family. Indeed, GH109_Pool2 is the first characterized β-N-acetylglucosaminidase/β-glucosidase of family GH109, expanding the activities repertoire of this CAZy family. Moreover, the GH109_Pool2 activity needed the addition of exogenous NAD^+ and the enzyme cleaved both α- and β-linked substrates, two characteristics only known so far in GH4 enzymes, suggesting structural and functional analogies never observed before between the GH4 and GH109 families.

2.3. Conclusion

The metagenomic study of Pisciarelli solfatara has allowed the discovery of several novel genes encoding for putative CAZymes; namely, 278 and 308 from Pool 1 and 2, respectively. After GTs, GHs represent the most abundant class of CAZymes identified (36 % and 38% for Pool1 and Pool2, respectively). Two novel GHs were biochemically characterized in detail: a novel GH5, subfamily 19 (GH5_19), β-mannanase/β-1,3-glucanase showed substrate specificity well correlating with the hemicelluloses present in the vegetation surrounding the sampling site. In addition, as part of my PhD thesis, the first NAD^+-dependent GH109 from Archaea (GH109_Pool2) was identified and characterized. The kinetic characterization of this enzyme showed features that differ significantly from the other known members of this family. Indeed, GH109_Pool2 is the first β-N-acetylglucosaminidase/β-glucosidase, expanding the activities repertoire of this CAZy family. In fact, although the total number of entries in GH109 increased to 1,265 (May 2021) still only 7 have been characterized. In addition, GH109_Pool2 was activated by exogenous NAD^+ addition, differently to the other characterized members of the family, in which the NAD^+ molecule was tightly bound to the enzymes in their recombinant form. This study demonstrated the power of the metagenomic approach to discover new enzyme sequences and to explore microbial biodiversity from extreme environments, which are still unexplored.

CHAPTER 3

3. Novel esterase from *G. thermodenitrificans*

3.1. Overview

Many known microbial genomes contain sequences encoding for enzymes that have never been characterized. In the framework of the discovery of novel enzymes from extremophiles, we found in the genome of the thermophilic microorganism *Geobacillus thermodenitrificans* a sequence encoding for a novel esterase. *In silico* analysis revealed that the enzyme belongs to the enterochelin/enterobactin esterase group of enzymes that are present only in siderophore microorganisms. These microorganisms synthetize and use siderophore molecules (i.e., enterochelin) to capture iron in the outer environment. Once the iron was caught, the complexes Fe-siderophore are imported inside the cell to be disassembled and free the metal. The hydrolysis of the complex is operated by a specific group of esterases, little characterized so far and known mainly for pathogenic bacteria. This sophisticated mechanism is suggested to play an essential role in the virulence of pathogens, indeed, the depletion of this iron-pathway cancels or drastically reduces the virulence of these microorganisms [138]. The characterization of the novel enterochelin/enterobactin esterase has revealed interesting potential applications. Moreover, the work shed light on this esterase group classification, which has never been recognized as a family in the lipase/esterase classification. In this chapter I will report the entire work in which my contribution has been predominant**.**

3.1.1. Geobacillus thermodenitrificans NG80-2

Geobacillus was recognized as a separated genus from *Bacillus* in 2001 by Nazina et al. [139]. Indeed, the species *G. thermodenitrificans* was classified for the first time in 2000 by Manachini et al. as *Bacillus thermodenitrificans* [140]. These microorganisms are thermophiles, facultative anaerobes, spore-forming, and using a wide range of carbon sources: aromatic compounds, alcohols, organic acids, and carbohydrates. They are ubiquitous microorganisms isolated from several marine and terrestrial thermal habitats. The genus is an exciting source for biocatalysis involved in biotechnological processes exhibiting a broad repertoire of hydrolytic enzymes, able to degrade complex polysaccharides and a broad range of alkanes [141]. *G. thermodenitrificans* NG80-2 was isolated in deep-subsurface oil reservoir in Dagang oilfield, Northern China. The microorganism has a temperature growth range between 45 °C and 73 °C and a pH range between 6.0 and 8.0. Moreover, it can grow in a minimal medium supplemented by crude oil and paraffin, showing the capability to degrade and oxidize long-chain alkane to ethanol [142].

Extremophiles
https://doi.org/10.1007/s00792-019-01093-9

ORIGINAL PAPER

Identification of a novel esterase from the thermophilic bacterium *Geobacillus thermodenitrificans* NG80-2

Nicola Curci[1,2] · Andrea Strazzulli[1,3] · Federica De Lise[1] · Roberta Iacono[1,2] · Luisa Maurelli[2] · Fabrizio Dal Piaz[4] · Beatrice Cobucci-Ponzano[2] · Marco Moracci[1,2,3]

Received: 19 February 2019 / Accepted: 16 April 2019

Abstract

In the framework of the discovery of new thermophilic enzymes of potential biotechnological interest, we embarked in the characterization of a new thermophilic esterase from the thermophilic bacterium *Geobacillus thermodenitrificans*. The phylogenetic analysis of the GTNG_0744 esterase indicated that the sequence belongs to the enterochelin/enterobactin esterase group, which have never been recognized as a family in the lipases/esterase classification. These enzymes catalyze the last step in the acquisition of environmental Fe^{3+} through siderophore hydrolysis. In silico analysis revealed, for the first time, that the machinery for the uptake of siderophores is present in *G. thermodenitrificans*. The purified recombinant enzyme, EstGtA3, showed different substrate specificity from known enterochelin/enterobactin esterases, recognizing short chain esters with a higher specificity constant for 4-NP caprylate. The enzyme does not require cofactors for its activity, is active in the pH range 7.0–8.5, has highest activity at 60 °C and is 100% stable when incubated for 16 h at 55 °C. DTT, β-mercaptoethanol and Triton X-100 have an activating effect on the enzymatic activity. Organic solvents have in general a negative effect on the enzyme, but n-hexane is a strong activator up to 150, making EstGtA3 a good candidate for applications in biotechnology.

Keywords Esterase · Bacterium · *Geobacillus* · Solvent tolerance · Thermostability

Abbreviations

4-NP	4-nitrophenyl
4-NPC12	4-nitrophenyl laurate
4-NPC2	4-nitrophenyl acetate
4-NPC4	4-nitrophenyl butyrate
4-NPC8	4-nitrophenyl caprylate
DTT	Dithiothreitol
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
EDTA	Ethylenediaminetetraacetic acid
EstGtA3	esterase from *G. thermodenitrificans* NG80-2
EtOH	Ethanol
MeOH	Methanol
MES	2-(N-morpholino)ethanesulfonic acid
PMSF	phenylmethane sulfonyl fluoride
SDS–PAGE	Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis
SDS	Sodium Dodecyl Sulphate
TCV	inhibitor tris-catechol vector

Communicated by H. Atomi.

Nicola Curci and Andrea Strazzulli contributed equally to the work

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00792-019-01093-9) contains supplementary material, which is available to authorized users.

✉ Beatrice Cobucci-Ponzano
beatrice.cobucciponzano@ibbr.cnr.it

[1] Department of Biology, University of Naples "Federico II", Complesso Universitario Di Monte S. Angelo, Via Cupa Nuova Cinthia 21, 80126 Naples, Italy

[2] Institute of Biosciences and BioResources, National Research Council of Italy, Via P. Castellino 111, 80131 Naples, Italy

[3] Task Force On Microbiome Studies, University of Naples Federico II, Naples, Italy

[4] Department of Medicine, Surgery and Dentistry, University of Salerno, Via Giovanni Paolo II, 84084 Fisciano, SA, Italy

Published online: 03 May 2019

Introduction

Geobacillus thermodenitrificans, formerly *Bacillus thermodenitrificans* (DSM 465^T and DSM 466 strains), is an aerobic or facultative anaerobic Gram-positive thermophilic bacterium belonging to the order of bacillales, which forms endospores (Manachini et al. 2000). Members of this genus, which are ubiquitous and have been isolated from thermal springs and waters of oil fields, include moderate thermophiles and neutrophiles using as source of energy a wide range of hydrocarbons such as aromatic compounds, alcohols, organic acids and carbohydrates (Nazina et al. 2001). The *G. thermodenitrificans* strain DSM 465^T grows in temperature and pH ranges of 40–70 °C and 6.0–8.0, respectively, and, though inhibited by sodium chloride > 5% (w/v), it tolerates phenol concentrations between 10 and 20 mM (Manachini et al. 2000). This strain is able, even under anaerobiosis, to reduce nitrates and nitrites in gaseous form and can use different carbohydrates as carbon sources, namely glucose, fructose, maltose, mannose, lactose, cellobiose, galactose, xylose, ribose and arabinose.

Another strain, identified as *G. thermodenitrificans* NG80-2, was isolated at a depth of about 2,000 meters and at 73 °C from the water of formation of a deep oil reserve, in Degang, in northern China (Wang et al. 2006). The optimal growth temperature and pH of NG80-2 strain are similar to the DSM 465 ^{T} strain and, recently, a β-xylosidase and a xylanase involved in the concerted degradation of xylan have been also identified in this organism (Jain et al. 2014; Marcolongo et al. 2015). However, this strain differs by the ability of oxidizing long-chain alkanes to alcohols. Indeed NG80-2 grows normally, though after a long latency phase, in a minimal medium supplemented with crude oil or paraffin, demonstrating the ability to use these compounds as the only carbon source. Wang and collaborators (Wang et al. 2006) suggested that the NG80-2 strain, like most of degrading n-alkanes bacteria, is able to produce emulsifying agents, not yet characterized, and that the latency period observed was no more than the time required to emulsify the long chains of alkanes.

All these characteristics make *G. thermodenitrificans* NG80-2 a promising source of new enzymes of biotechnological interest. Lipases and esterases from this organism can be used for the hydrolysis and modifications of fatty acids longer or shorter than 12 carbon atoms in food and paper industries, in bioremediation processes, for the production of biodiesel and for the synthesis of optically pure compounds of agrochemical, pharmaceutical, and cosmetic interest. In addition, glycosidases and carbohydrate esterases have useful applications in the degradation of lignocellulose for the production of second-generation bioethanol (Bornscheuer 2002; Cobucci-Ponzano et al. 2015; Jaeger and Eggert 2002; Kumagai et al. 2018).

Lipases and esterases, in aqueous solutions, catalyze the cleavage of an ester bond between a carboxylic acid and an alcohol, while, in organic solvents or in non-aqueous media, reverse hydrolysis or transesterification reactions become possible (Levisson et al. 2007). These enzymes, though showing limited sequence homology, have highly conserved tertiary structure with a typical α/β hydrolase fold composed of eight parallel β-sheets connected by six α-helices. The active site is located within an α-helix and is composed of a catalytic triad of Ser, Asp/Glu, and His, with the serine widely conserved in the pentapeptide Gly-X-Ser-X-Gly (López-López et al. 2014). The microbial lipases/esterases classification is based on the amino acid sequence producing eight families (I–VIII) (Arpigny and Jaeger 1999), which were subsequently integrated over time with new families (up to XVIII), some of them discovered only recently by metagenomic approaches (López-López et al. 2014; Samoylova et al. 2018). In addition, although the presence of the conserved Gly-X-Ser-X-Gly motif was a classical trademark of lipases/esterases in protein databases, it is worth noting that several of them lacked this pentapeptide making the sequence-based classification ineffective without a detailed biochemical characterization.

So far, two esterases from two strains of *G. thermodenitrificans* have been already identified and characterized; EstGtA2 from strain CMB-A2 and EstL5 from strain T2, belonging to families XV and VIII, respectively, of the lipases/esterases classification (Charbonneau et al. 2010; Yang et al. 2013).

In the framework of our study of new thermophilic enzymes of potential biotechnological interest (Cobucci-Ponzano et al. 2010a, b, 2011, 2012; Iacono et al. 2019; Sirec et al. 2012; Strazzulli et al. 2017a), we embarked in the search in the genome of *G. thermodenitrificans* strain NG80-2 of genes encoding hypothetical lipases/esterases. Here, we report the cloning, expression, and characterization of a new thermophilic esterase that allowed to define in the lipases/esterases classification a new family not assigned to date.

Materials and methods

Materials

One Shot® TOP10 Competent Cells (Invitrogen), BL21 Star™ (DE3) *E. coli* (Invitrogen) and Plasmid pET/101/D-TOPO (Invitrogen) were used for sub-cloning and expression.

Cloning of GTNG_0744

The open reading frame GTNG_0744 encoding for a hypothetical esterase was amplified from genomic DNA of *G. thermodenitrificans* NG80-2 using 0.2 µM of the forward (5′-TAATACGACTCACTATAGGG-3′) and reverse (5′-TATGCTAGTTATTGCTCAG-3′) primers in a 50-µL reaction mix containing 0.2 µM dNTP, 900 ng of the template and 2.5 U of *Pfu* Ultra HF Polymerase (Agilent). The PCR program was conducted as follows: 95 °C for 5 min; 5 cycles (95 °C for 1 min, 50 °C for 1 min, 72 °C for 1.5 min); 30 cycles (95 °C for 1 min, 60 °C for 1 min, 72 °C for 1.5 min); 72 °C for 10 min. The PCR product was purified by Quantum Prep® PCR kleen Spin Columns kit (Bio-Rad) and cloned in pET/101/D-TOPO plasmid by Champion™ pET101 Directional TOPO™ Expression kit (Invitrogen by Thermo Fischer Scientific). The resulting plasmid, pET101TOPO-EstGTA3, was used to transform *E. coli* One Shot® TOP10 and the positive clones were selected by colony PCR.

Phylogenetic analysis and homology modeling

The deduced amino acid sequence of GTNG_0744 was multialigned with the sequences of the esterase families as reported in (Samoylova et al. 2018) using Clustal Omega (Sievers et al. 2011).

A phylogenetic tree was made using MEGA 7.0 (Kumar et al. 2016) with the maximum likelihood method using settings of Poisson Model, homogenous patterning between lineages and a bootstrap consensus from 1000 replicates.

Homology modeling of EstGtA3 structure was performed by SWISS-MODEL server (Arnold et al. 2006) using the ferric enterobactin esterase PfeE mutant S157A from *Pseudomonas aeruginosa* (PDB entry 6gi1) as a template. Stereochemical quality of the model was analyzed by the PROCHECK program (Laskowski et al. 1993) and PyMol 1.0 was used to analyze and visualize the structure.

Expression and purification

Escherichia coli BL21 Star™ (DE3) One Shot™ competent cells were transformed with the plasmid pET-101TOPO-EstGTA3. An overnight culture was inoculated into 4 L of LB and grown at 37 °C until OD_{600} of 0.6–0.8. The expression of the *estgta3* gene was induced by addition of 1 mM IPTG at 37 °C for 16 h. The cellular pellet was harvested by centrifugation at 5,000 × g and resuspended 1:5 (g:mL) in Tris–HCl 50 mM, NaCl 150 mM pH 7.5 in the presence of 25 mg of lysozyme, 5 mg of DNase I and 1% Triton X-100. After incubation of 1 h at room temperature, 5 cycles of freeze–thawing were performed and the resulting lysate was centrifuged at 16,000×*g* for 30 min.

The supernatant was diluted 1:2 (vol:vol) with Tris–HCl 50 mM, NaCl 150 mM pH 7.5 (buffer A) and loaded on an AKTA Explorer FPLC system (GE Healthcare) equipped with a His Trap HP crude column (1 mL) previously equilibrated with the same buffer. The elution was performed with buffer B (Buffer A supplemented with imidazole 500 mM) as follows: 20 mL 100%A, 20 mL at 5%B, 20 mL at 50%B and 20 mL 100%B. The relevant fractions were pooled and dialyzed against buffer A.

The dialyzed sample was incubated at 55 °C for 30 min and centrifuged at 16,000×*g* for 30 min. The supernatant was loaded on ion metal affinity bench column of kit Protino® Ni-TED 1000 Packed Columns (Machery-Nagel), equilibrated with Tris–HCl 50 mM NaCl 150 mM pH 7.0. The protein was eluted with 2 mL buffer B and dialyzed against buffer A.

The resulting sample was concentrated by ultrafiltration on Amicon® Ultra-4 centrifugal filter (cut-off 10,000 Da) (Merk Millipore) and loaded on to Superdex™ 75 10/300 GL (GE Healthcare) in buffer A at 0.7 mL/min for 36 mL.

Protein concentration was determined by Bradford method with bovine serum albumin (BSA) as standard (Bradford 1976). Native molecular mass was examined by gel filtration calibrated with β-amylase (200,103 Da), BSA (66,103 Da), carbonic anhydrase (29,000 Da), cytochrome c (12,400 Da) and aprotinin (6500 Da).

The identity of the protein was confirmed using a mass spectrometry-based approach; briefly, purified protein was analyzed by SDS–PAGE and the resulting bands underwent in-gel trypsin digestion. The obtained peptides were analyzed by LC–MS/MS using an Orbitrap XL instrument (Thermo Fisher, Waltham, MA, USA) equipped with a nano-ESI source coupled with a nano-Acquity capillary UPLC (Waters, Milford, MA, USA). Peptides were separated with a capillary BEH C18 column (0.075 × 100 mm, 1.7 µM, Waters) using aqueous 0.1% formic acid (A) and CH_3CN containing 0.1% formic acid (B) as mobile phases. Peptides were eluted by means of a linear gradient from 5 to 50% of B in 90 min, at a 300 nL/min flow rate. Mass spectra were acquired over an *m/z* range from 400 to 1800. To achieve protein identification, MS and MS/MS data underwent Mascot Search Engine software analysis to interrogate the National Center for Biotechnology Information nonredundant (NCBInr) protein database. Parameters sets were: trypsin cleavage; carbamidomethylation of cysteines as a fixed modification and methionine oxidation as a variable modification; a maximum of two missed cleavages; false discovery rate, calculated by searching the decoy database, 0.05.

Enzyme activity assay

The esterase activity on 4-nitrophenyl (4-NP) esters was determined by measuring the amount of 4-nitrophenol released after the addition of the enzyme. The 4-NP esters were dissolved in acetonitrile and the molar extinction coefficient of 4-NP at 405 nm in Tris–HCl pH 7.5 (13.1 M^{-1} cm^{-1}) was used for the determination of 4-NP concentration in solution.

The standard assay was performed at 55 °C in a final volume of 500 µL of a reaction mixture containing 50 mM Tris–HCl pH 7.5, 4% propan-2-ol, 2 mM 4-NP acetate and 2 µg of purified enzyme. The production of 4-NP was monitored continuously at 405 nm using a double-beam spectrophotometer with a thermal control unit. The background hydrolysis of the substrate was subtracted using a reference sample identical to the reaction mixture, but without enzyme. One unit (U) of esterase activity was defined as the amount of enzyme that released 1 µmol of 4-nitrophenol per min at the conditions described.

Substrate specificity

The substrate specificity of the enzyme was determined on 4-NP esters with different length of acyl chain in standard conditions, namely: 4-NP acetate (4-NPC2), 4-NP butyrate (4-NPC4), 4-NP caprylate (4-NPC8), and 4-NP laurate (4-NPC12). Steady-state kinetic constants were measured in standard conditions using concentrations in the range 0.25–4.0 mM for 4-NPC2 and 4-NPC4, and 1.9–124 µM for 4-NPC8. The data were analyzed by nonlinear regression using the Michaelis–Menten equation using Prism 5.

Effect of pH and temperature on the enzyme activity

The optimal pH of esterase activity was determined at 55 °C with 4-NPC2 as substrate, over the range pH 6.0–9.0 in the following buffers (50 mM final concentration): NaOH-MES (pH 6.0–pH 7.0), NaOH-HEPES (pH 7.0–7.5), and Tris–HCl (pH 7.5–9.0).

The temperature dependence of the activity of the enzyme was determined in the range 30–75 °C in 50 mM Tris–HCl (pH 7.5) buffer with 4-NPC2 as substrate.

The thermostability was evaluated by incubating the enzyme in 50 mM Tris–HCl buffer (pH 7.5) at 55 °C and 65 °C, up to 16 h. Aliquots (2 µg) were withdrawn at the indicated times and assayed on 4-NPC2 as substrate at standard conditions. The specific activity of the enzyme before incubation ($s = 0$ min) was considered 100%.

Effect of metal salt, inhibitors and detergents on the enzyme stability

The effect of metal salts (NaCl, KCl, $CaCl_2$, $FeSO_4$, $MgSO_4$, $CuCl_2$, $ZnSO_4$, $MnSO_4$), inhibitors and detergents (EDTA, DTT, SDS, Triton X-100, PMSF, β-mercaptoethanol) on the enzyme stability was determined by adding these reagents at the final concentration of 1 or 10 mM to the enzyme solution.

The enzyme solution was pre-incubated for 1 h at 55 °C in the presence of the indicated metal salts, inhibitors or detergents. After incubation, aliquots of the enzyme solution, containing 2 µg of EstGtA3, were used to measure the esterase activity on 4-NPC2 at standard conditions. The specific activity without reagent was measured after the same pre-incubation and was defined as 100%.

Effect of organic solvents on the enzyme stability

The effect of organic solvents on the esterase activity was examined by adding methanol, ethanol, n-hexane, acetone, n-butanol, DMSO, chloroform, propan-2-ol, acetonitrile, and DMF at the final concentration of 10 or 30% to the enzyme solution. The enzyme solution was pre-incubated with each indicated solvent for 1 h at 55 °C. After incubation, aliquots of the enzyme solution, containing 2 µg of EstGtA3, were used to measure the esterase activity on 4-NPC2 at standard conditions. The specific activity measured using an enzyme solution incubated for 1 h at 55 °C without solvents was defined as 100%.

Results

In silico analysis of the esterase EstGtA3

Sequence analysis of the genome of the thermophilic bacterium *Geobacillus thermodenitrificans*, strain NG80-2 (Wang et al. 2006) revealed the presence of the open reading frame GTNG_0744 that encodes for a protein of 254 amino acids with an estimated molecular weight of 26,000 Da (accession: WP_008878924). A preliminary analysis of GTNG_0744 by BlastP against NR database, excluding *Geobacillus* (taxon ID:129,337), revealed identities between 90 and 57% with several hypothetical esterases belonging to *Anoxybacillus spp and Bacillus spp,* including an uncharacterized protein from *Flavobacterium thermophilum* annotated as enterobactin esterase (89% identity) (Table S1).

To determine to which family of esterases GTNG_0744 belongs, the sequence was aligned by ClustalO against the representative sequences of the families previously identified (Ranjan et al. 2018; Samoylova et al. 2018) (Fig. 1). The phylogenetic analysis clearly indicates that GTNG_0744

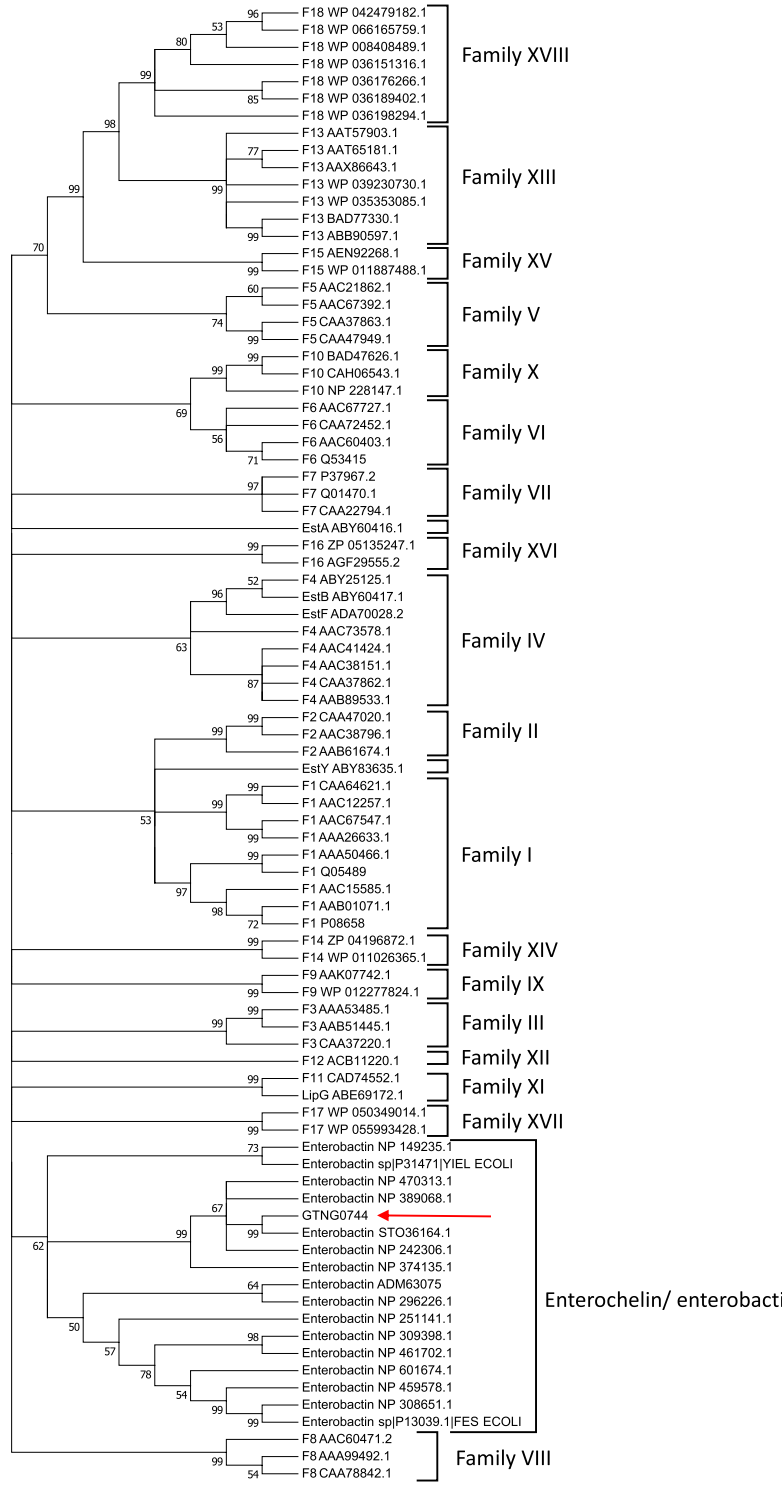


Fig. 1 Phylogenetic tree of GTNG_0744 (EstGtA3). Analysis was performed with the maximum likelihood and a bootstrap consensus from 1000 replicates. Bootstrap values are indicated at respective nodes. The arrow indicates GTNG_0744 (EstGtA3) from *G. thermodenitrificans* NG80-2

belongs to a group of enterochelin/enterobactin esterases branching close to families VIII and XVII, both grouping enzymes with similarity to the class of β-lactamases (Castilla et al. 2017; López-López et al. 2014; Ramnath et al. 2017).

The inspection of GTNG_0744 amino acidic sequence showed that this novel esterase contains the conserved GDSLG motif, which is different from any other lipase/esterase family currently classified, but it is present in all the sequences encoding for enterobactin esterases (Ranjan et al. 2018). This particular group of esterases is involved in the acquisition of environmental Fe^{3+} by hexadentate catecholate siderophores, high-affinity ferric ion specific chelators excreted under iron starvation by various microorganisms like bacteria, fungi and some plants. In conditions of iron limitation, bacteria produce hexadentate catecholate siderophores by non-ribosomal peptide synthesis, which differs between Gram-positive and Gram-negative bacteria (Khan et al. 2018). Once the siderophores bind environmental Fe^{3+}, they get internalized through an active transport machinery (Fig. 2) and the intracellular iron release is promoted by the enterobactin esterase hydrolysis of the hexadentate catecholate scaffold (Abergel et al. 2009; Lin et al. 2005).

The search for the amino acid sequences homologous to those encoding for enzymes involved in the iron uptake of siderophores in *G. thermodenitrificans* revealed sequences homologous to those encoding for the proteins involved in the uptake machinery, with identity between 28 and 75% and similarity between 22 and 86% (Table 1). The genomic organization of these ORFs is similar to the gene clusters FepBCD and FeuABC in *E. coli* and *B. subtilis*, respectively (Fig. S1) (Chenault and Earhart 1991; Ollinger et al. 2006). These evidences, along with the identification of GTNG_0744, strongly suggest that *G. thermodenitrificans* could be a siderophore microorganism and that product of GTNG_0744 could be possibly involved in siderophore hydrolysis.

The homology model of GTNG_0744 structure (QMEAN -3.83, overlay RMS 0.581 Å) was obtained using SWISS-MODEL server (https://swissmodel.expasy.org) and the ferric enterobactin esterase PfeE mutant S157A from *Pseudomonas aeruginosa* as a template (PDB 6gi1: 19% identity, 27.8% similarity) (Perraud et al. 2018). The analysis revealed that GTNG_0744 shows a classical α/β hydrolase fold with 8 β -sheet and 9 α-helices (Fig. 3 a and b). The validation by PROCHECK (Laskowski et al. 1993) indicates that the model obtained has 96.2% residues in the most favored regions of Ramachandran plot, 1.9% residues in the generously allowed regions and 1.9% residues disaligned. Moreover, the superimposition (RMS 0.656 Å) of the GTNG_0744 model with PfeE wild type (PDB 6gi5) in complex with the competitive inhibitor tris-catechol vector (TCV) (Perraud et al. 2018) revealed also a conserved spatial disposition of the pentapeptide GXSXG and, in particular, the orientation of the putative catalytic Ser129 (Fig. 3c, d). Consistent with these results, we named this gene *estgta3* and the codified protein EstGtA3. Both sequence analysis and the in silico homology model suggest that this enzyme may seed a new lipases/esterase family, thus we embarked in the biochemical characterization.

Expression and purification of EstGtA3

The coding sequence of EstGtA3 esterase was cloned into pET/101/D-TOPO vector and expressed in *E. coli* BL21 (DE3). The recombinant protein was purified by a Ni–NTA affinity chromatography, a heat-fractionation step at 55 °C, a second Ni–NTA affinity and a gel filtration chromatography. The purification procedure led to a 90% pure protein by SDS–PAGE with a final yield of 7%.

The molecular mass of the purified enzyme on SDS–PAGE is about 30 kDa as predicted for the fusion protein EstGtA3 with the C-terminal 6-His tag (Fig. S2). The molecular mass of EstGtA3 in native conditions, determined by gel filtration chromatography, is 26,000 Da, which is in good agreement with that calculated from the amino acid sequence (25,940 Da) (Fig. S3), indicating that EstGtA3 is a monomer in solution. The identity of the protein was further confirmed using a mass spectrometry-based approach (Table S2).

Substrate specificity of EstGtA3

The substrate specificity of EstGtA3 was determined by assaying the enzyme on 4-nitrophenyl esters, namely, 4-nitrophenyl acetate (4-NPC2), 4-nitrophenyl butyrate (4-NPC4), 4-nitrophenyl caprylate (4-NPC8), and 4-nitrophenyl laurate (4-NPC12). The assays were conducted in 50 mM Tris–HCl pH 7.5, 4% propan-2-ol, at 55 °C as detailed in Materials and Methods.

The assays of EstGtA3 on 4-NP esters revealed that the enzyme exhibits highest activity on short-chain fatty acids (4-NPC2 > 4-NPC4 > 4-NPC8) and no activity was observed on 4-NPC12. The esterase followed a typical Michaelis–Menten behavior for the hydrolysis of the substrates and the steady-state kinetic constants were determined (Table 2). The highest specificity constant on 4-NPC8 is explained by the micromolar K_M on this substrate; indeed, the k_{cat} was 5.5-fold lower than on 4-NPC2.

The bioinformatic analysis of EstGtA3 showed that this novel esterase groups with enterochelin/enterobactin esterases (Fig. 1), possibly forming a new family (Ranjan et al. 2018). To test if EstGtA3 could hydrolyze siderophores, the recombinant enzyme was assayed on commercial bacillibactin (2,3-dihydroxybenzoate-glycine-threonine trimeric ester) and enterobactin

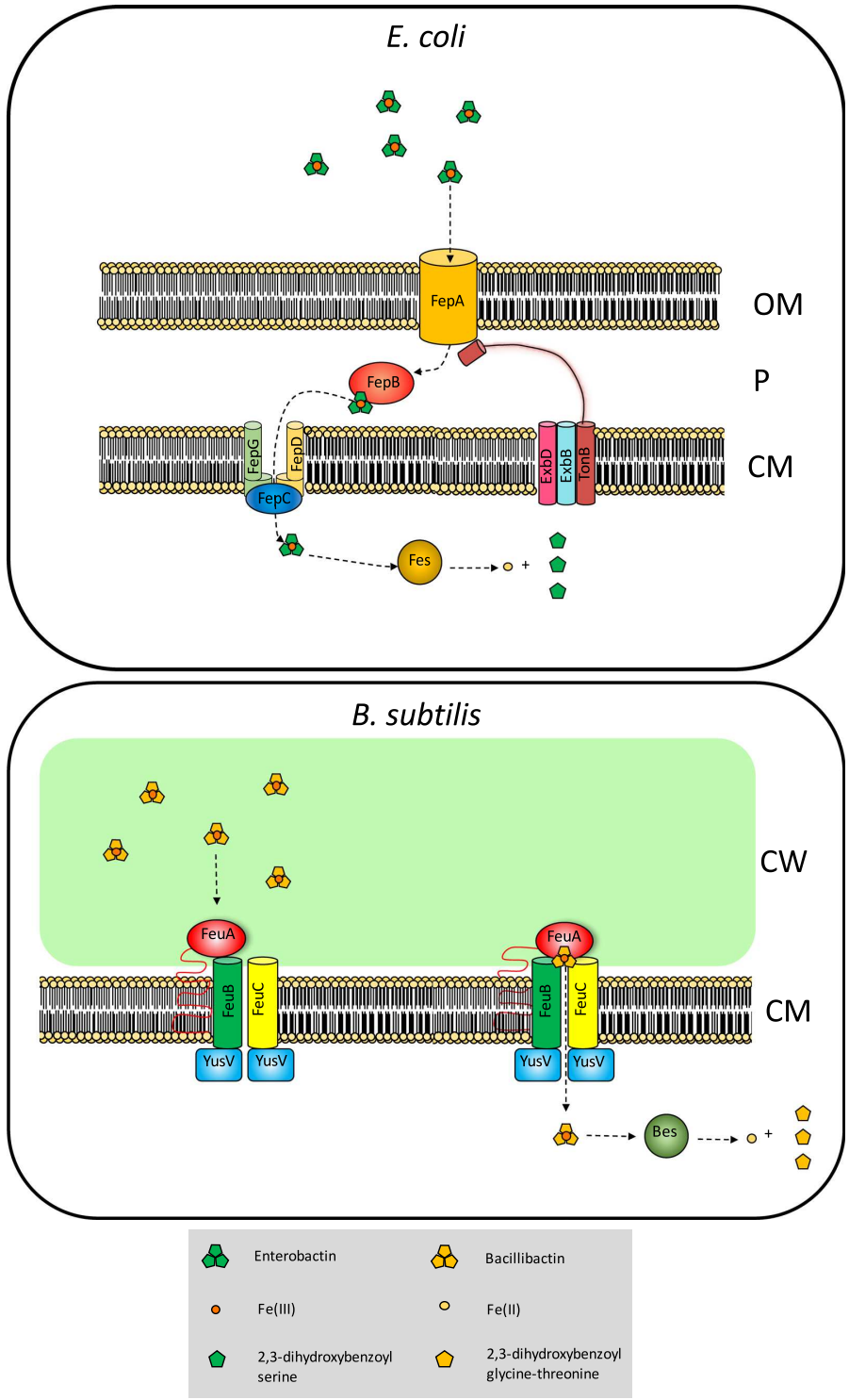


Fig. 2 Iron acquisition machinery used by *E.coli* and *B. subtilis*. (OM) outer membrane (P) periplasm (CM) cytosolic membrane (CW) cell wall

Table 1 Orfs of *G. thermodenitrificans* homologous to the proteins involved in the iron acquisition systems used by Gram-positive and negative bacteria

Gram	Protein name	*G. thermodenitrificans* orf	Description	% Identity	% Similarity
+	FeuA	GTNG_1317	ABC transporter periplasmic binding domain	28	46.7
+	FeuB	GTNG_0173	ABC transporter, permease protein, BtuC-like	36	59.2
+	FeuC	GTNG_0172	ABC transporter, permease protein, BtuC-like	35	55.5
+	YusV	GTNG_1320	ATP-binding domain of ABC transporters	75	86.5
−	FepB	GTNG_1317	ABC transporter periplasmic binding domain	24	22.3
−	FepC	GTNG_1320	ABC transporter-like with AAA + ATPase domain	52	69.3
−	FepD	GTNG_1318	ABC transporter, permease protein, BtuC-like	44	51.8
−	FepG	GTNG_1319	ABC transporter, permease protein, BtuC-like	39	56.2

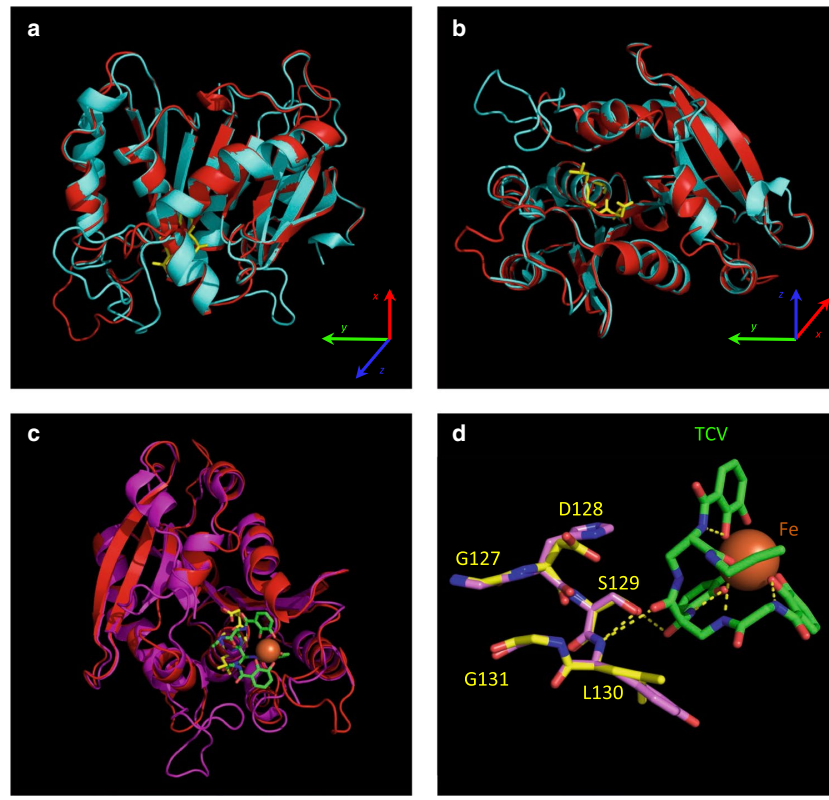


Fig. 3 3D model structure of EstGtA3. **a**, **b** Superimposition of EstGtA3 model (red) with the 3D structure of PfeE enterobactin esterase mutant S157A from *P. aeruginosa* (cyano, PDB 6gi1). The EstGtA3 catalytic pentapeptide is colored in yellow. **c** Superimposition of EstGtA3 model (red) with PfeE wild type (magenta, PDB 6gi5) in complex with the triscatechol vector (TCV, green). **d** Close-up of the pentapeptide of EstGtA (yellow) and PfeE wild type (pink) in complex with TCV

Table 2 Steady-state kinetic constants of EstGtA3

Substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($mM^{-1}\ s^{-1}$)
4-NPC2	0.97 ± 0.18	6.80 ± 0.50	7.01
4-NPC4	0.64 ± 0.18	3.03 ± 0.25	4.73
4-NPC8	0.005 ± 0.001	1.24 ± 0.04	248

(2,3-dihydroxybenzoate-L-serine trimeric ester) which are the siderophores from Gram-positive and Gram-negative bacteria, respectively (Fig. 2). No hydrolytic activity could be observed (Fig S4); possibly, this is due to the high specificity of this class of enzymes toward their substrates (Abergel et al. 2009). Therefore, the identification of a different siderophore in *Geobacillus* needs further studies.

Effect of pH and temperature on the esterase activity of EstGtA3

The activity of EstGtA3 was measured at different temperatures and displayed the highest hydrolytic rates at 60 °C; EstGtA3 was 100% stable when incubated for 16 h at 55 °C and assayed at standard conditions, while only about 20% residual activity was observed after 5 min at 65 °C (Fig. 4a, b). The enzyme showed the highest activity in the pH range 7.0–8.5 (Fig. 4c).

Effect of metals, inhibitors and detergents on EstGtA3

The effect of metal ions, inhibitors and detergents on EstGtA3 activity was analyzed after incubation of the enzyme with 1 mM or 10 mM of each metal, inhibitor or detergent, for 1 h at 55 °C (Fig. 5a, b). The specific activity of the enzyme was reduced to 50% by 10 mM concentrations of $FeSO_4$, $CuCl_2$, $ZnSO_4$, and KCl (Fig. 5a); however, EDTA had no inhibiting effect (Fig. 5b). SDS and PMSF inhibited EstGtA3 at both concentrations; instead, remarkably, DTT, 2-mercaptoethanol, and Triton X100 showed an activating effect (Fig. 5b).

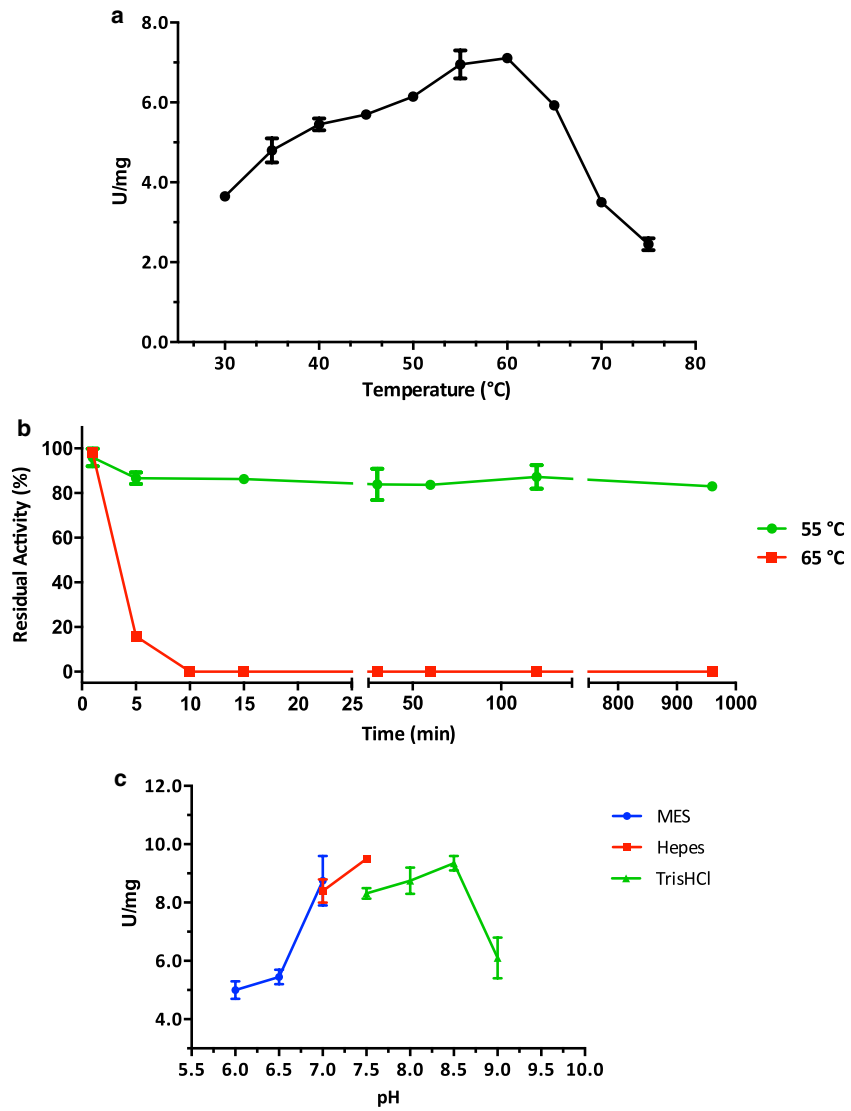


Fig. 4 Effect of pH and temperature on the EstGtA3 activity. **a** The effect of temperature on the activity of the enzyme was measured at standard conditions by using 4-NPC2 substrate at the indicated temperatures. **b** The effect of temperature on the stability of EstGtA3 was measured by incubating the enzyme in 50 mM Tris–HCl (pH 7.5) buffer at 55 °C and 65 °C up to 16 h; residual activity was measured at standard conditions on 4-NPC2; 100% was the specific activity before incubation. **c** The effect of pH on the activity of the enzyme was measured using 4-NPC2 as a substrate at 55 °C in the indicated buffers

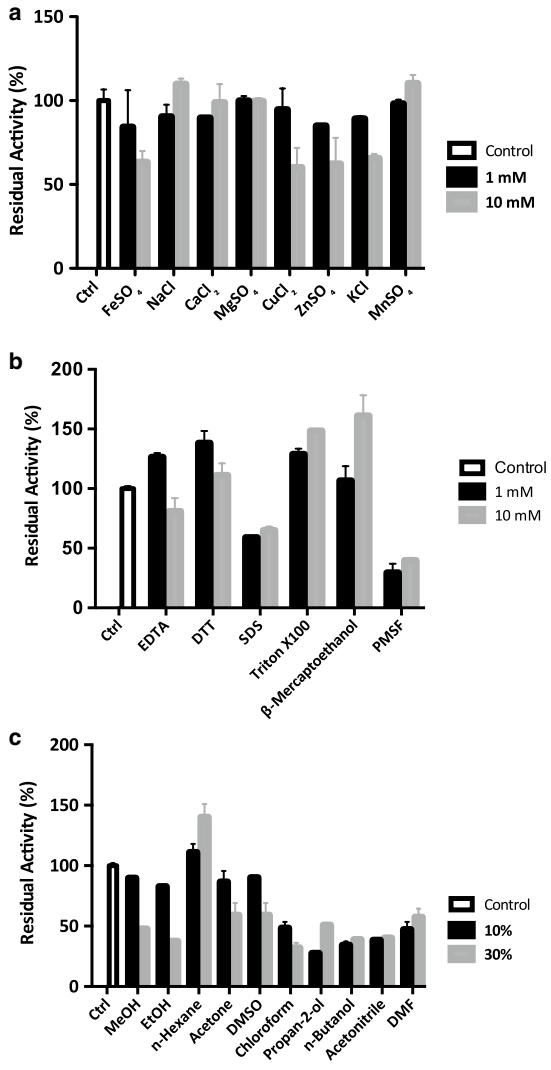


Fig. 5 Effect of metal ions, inhibitors, detergents, and organic solvents on the EstGtA3 stability. Standard esterase assays were performed in the presence of 1 mM and 10 mM concentrations of metal ions (**a**), inhibitors and detergents (**b**). The effect of organic solvents on the enzyme stability (**c**) was measured by incubating the enzyme in the presence of 10% and 30% (v/v) concentrations. In all cases, EstGtA3 was assayed at 55 °C, in standard conditions, by using 4-NPC2 substrate

Effect of organic solvents on EstGtA3

The effect of organic solvents on EstGtA3 was evaluated after incubation of the enzyme in the presence of solvents (10 and 30% v/v) for 1 h at 55 °C (Fig. 5c). Miscible organic solvents, namely, methanol, ethanol, acetone, DMSO, propan-2-ol, acetonitrile, and DMF had detrimental effect on the esterase activity, especially at 30% concentration. Among immiscible solvents, chloroform and n-butanol had detrimental effect on the esterase activity at both concentrations; while, interestingly, 30% n-hexane activated EstGtA3 up to 150% (Fig. 5c).

Discussion

(Hyper)thermophiles have long been a source of novel biocatalysts which can be used in several biotechnological applications (Iacono et al. 2016; Strazzulli et al. 2017b, c). In particular, thermally stable esterases can be applied in different food and pharmaceutical industrial processes as well as in the fine chemical industry and in biorefinery (Elleuche et al. 2015; Levisson et al. 2009). In this study, we focused on the identification, classification, and biochemical characterization of a recombinant novel esterase from the thermophilic bacterium *G. thermodenitrificans*.

The bioinformatic analysis of EstGtA3 revealed that this enzyme groups with enterochelin/ enterobactin enzymes which have not been yet classified into a family of lipase/ esterase classification. These enzymes are involved in the acquisition of environmental iron Fe^{3+} by hydrolysing intracellularly the hexadentate catecholate siderophores imported by specific machineries which differ in Gram-positive and Gram-negative bacteria (Fig. 2) (Khan et al. 2018). We showed for the first time that hypothetical proteins possibly involved in the iron uptake are present in *G. thermodenitrificans* and that their genes have an organization similar to those of *E. coli* and *B. subtilis* (Fig. S1) (Chenault and Earhart 1991; Ollinger et al. 2006). These data strongly suggest that *G. thermodenitrificans* could be a siderophore microorganism. However, EstGtA3 assayed on commercial bacillibactin and enterobactin, as example of siderophores from Gram-positive and Gram-negative bacteria, respectively, did not show hydrolytic activity (data not shown). This result is not unexpected as this class of enzymes shows high specificity toward their substrates (Abergel et al. 2009) and suggest the presence of a different siderophore in *Geobacillus* which is still unknown.

To date, ferric enterobactin esterases are still poorly characterized. Recently, sequences assigned to hypothetical enterobactin esterases have been identified also in a hot spring metagenomic sample collected in Khir Ganga (Himachal Pradesh, India) (Ranjan et al. 2018), but no thermophilic enterobactin esterases have been characterized into detail, leaving their biotechnological potential still unexploited.

The enzymatic characterization of EstGtA3 revealed a canonical esterase behavior. The enzyme is active in the pH range of 7.0–8.5. (Fig. 4c, Table 3), stable in solvents and surfactant and does not require cofactors for its activity.

Table 3 Comparison of characterized esterases identified in thermophilic microorganisms and from hot springs metagenomic.

Esterase	Source	Substrate: $\%k_{cat}/K_M$	pH Range	T optimum (°C)	References
EstGtA3	*Geobacillus thermodenitrificans NG80-2*	4-NPC2: 2.8 4-NPC4: 1.9 4-NPC8: 100	7.0–8.5	60	This work
EstL5	*Geobacillus thermodenitrificans T2*	4-NPC2: 24.7 4-NPC4: 100 4-NPC8: 8.4	7.0–9.0	60	Yang et al. (2013)
estUT1	*Ureibacillus thermosphaericus*	4-NPC2: 100 4-NPC4: 28.7 4-NPC8: 10.7	6.0–9.0	70–80	Samoylova et al. (2018)
EstEP16	*Metagenomic analysis*	4-NPC2: 100 4-NPC4: 98 4-NPC8: 2	7.5–9.5	60	Zhu et al. (2013)
MLC3	*Metagenomic analysis*	4-NPC2: 25 4-NPC4: 100 4-NPC8: 2	7.5–8.5	50	Ranjan et al. (2018)
SLC5	*Metagenomic analysis*	4-NPC2: 70 4-NPC4: 100 4-NPC8: –	8.0–9.0	35	Ranjan et al. (2018)
EstSp	*Metagenomic analysis*	4-NPC2: 100 4-NPC4: 16 4-NPC8: 3.7	8.0–9.0	40	Jayanath et al. (2018)
G.sp JM6 est	*Geobacillus sp. JM6*	4-NPC2: 88.5 4-NPC4: 100 4-NPC8: –	6.0–12.0	60	Zhu et al. (2015)
Est8	*Metagenomic analysis*	4-NPC2: 100 4-NPC4: 6 4-NPC8: –	8.0–9.0	30	Pereira et al. (2017)
LacH	*Brevundimonas sp. LY-2*	4-NPC2: 100 4-NPC4: 80 4-NPC8: 4.5	6.0–7.0	40	Zhang et al. (2017)

EstGtA3 shows the highest activity on short-chain fats, revealing, however, a higher substrate specificity on 4-NPC8. This last feature is uncommon in esterases with high activity on 4-NPC2 and 4-NPC4, which show very low specificity, or no activity, on 4-NPC8 (Table 3). Moreover, esterases active on 4-NPC8 are generally active also on 4-NPC12 (Li et al. 2018; Yu et al. 2018); while, surprisingly, no activity was detected for EstGtA3 on this latter substrate.

EstGtA3 has optimal activity in the pH range of 7.0–8.5 and at 60 °C, showing 100% of stability at 55° C for 16 h, features comparable to other thermostable esterases, for example, EstL5 and EstEP16, from *G. thermodenitrificans* T2 and an unknown species from a metagenomic sample, respectively (Table 3).

The activity of EstGtA3 was not affected by detergents, inhibitors and metals at 1 mM concentration, while at 10 mM of $FeSO_4$, $CuCl_2$, $ZnSO_4$, and KCl, the specific activity is reduced up to 50%. In addition, EDTA did not affect the activity indicating that the enzyme does not have a structural metal ion. Instead, Triton X100 and molecules with thiol groups, such as DTT and 2-mercaptoethanol, showed an activating effect at both concentrations. By contrast, SDS and PMSF strongly reduced the activity of the enzyme at both concentrations tested. The effect of PMSF suggests that a serine residue might be involved in the catalytic site of the EstGtA3 like for other esterases (Samoylova et al. 2018). Organic solvents showed a general negative effect on the stability of EstGtA3, especially for immiscible solvents with a loss of activity up to 70%. By contrast, n-hexane has a clear activating effect on EstGtA3, as observed for other esterases (Kim et al. 2015; Li and Yu 2014; De Santi et al. 2014). This latter feature makes EstGtA3 interesting for the resolution of racemic mixtures and synthesis of optically pure compounds (Huang et al. 2016; Kanamori et al. 2005). The stability of the novel thermostable EstGtA3 to various chemical compounds and the lack of requirement for cofactors make this

enzyme very interesting for its exploitation in industrial processes, such as in laundry and washing sector, where the use of lipolytic enzymes has long been consolidated (Elleuche et al. 2015; Yang et al. 2018).

Acknowledgements We thank Francesco La Cara and collaborators at the Research Institute on Terrestrial Ecosystems (IRET) from the National Research Council of Italy for the gift of *Geobacillus thermodenitrificans* NG80-2 genome. We are grateful to Chiara Nobile and Marco Petruzziello at the Institute of Biosciences and BioResources (IBBR) from the National Research Council of Italy for administrative and technical assistance. This work was supported by a grant from the Italian Ministry of Research (MIUR) PON03PE_00107_1 BIOPOLIS.

References

Abergel RJ, Zawadzka AM, Hoette TM, Raymond KN (2009) Enzymatic hydrolysis of trilactone siderophores: where chiral recognition occurs in enterobactin and bacillibactin iron transport. J Am Chem Soc 131:12682–12692. https://doi.org/10.1021/ja903051q

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22:195–201. https://doi.org/10.1093/bioinformatics/bti770

Arpigny JL, Jaeger KE (1999) Bacterial lipolytic enzymes: classification and properties. Biochem J 343(Pt 1):177–183

Bornscheuer UT (2002) Microbial carboxyl esterases: classification, properties and application in biocatalysis. Fems Microbiol Rev 26:73–81. https://doi.org/10.1111/j.1574-6976.2002.tb00599.x

Bradford MM (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. https://doi.org/10.1016/0003-2697(76)90527-3

Castilla A, Panizza P, Rodriguez D, Bonino L, Diaz P, Irazoqui G, Rodriguez Giordano S (2017) A novel thermophilic and halophilic esterase from Janibacter sp. R02, the first member of a new lipase family (Family XVII). Enzyme Microb Technol 98:86–95. https://doi.org/10.1016/j.enzmictec.2016.12.010

Charbonneau DM, Meddeb-Mouelhi F, Beauregard M (2010) A novel thermostable carboxylesterase from *Geobacillus thermodenitrificans*: evidence for a new carboxylesterase family. J Biochem 148:299–308. https://doi.org/10.1093/jb/mvq064

Chenault SS, Earhart CF (1991) Organization of genes encoding membrane proteins of the *Escherichia coli* ferrienterobactin permease. Mol Microbiol 5:1405–1413

Cobucci-Ponzano B, Aurilia V, Riccio G, Henrissat B, Coutinho PM, Strazzulli A, Padula A, Corsaro MM, Pieretti G, Pocsfalvi G, Fiume I, Cannio R, Rossi M, Moracci M (2010) A New Archaeal beta-Glycosidase from *Sulfolobus solfataricus* seeding a novel retaining beta-glycan-specific glycoside hydrolase family along with the human non-lysosomal glucosylceramidase gba2. J Biol Chem 285:20691–20703. https://doi.org/10.1074/jbc.M109.086470

Cobucci-Ponzano B, Conte F, Strazzulli A, Capasso C, Fiume I, Pocsfalvi G, Rossi M, Moracci M (2010) The molecular characterization of a novel GH38 alpha-mannosidase from the crenarchaeon *Sulfolobus solfataricus* revealed its ability in de-mannosylating glycoproteins. Biochimie 92:1895–1907. https://doi.org/10.1016/j.biochi.2010.07.016

Cobucci-Ponzano B, Zorzetti C, Strazzulli A, Carillo S, Bedini E, Corsaro MM, Comfort DA, Kelly RM, Rossi M, Moracci M (2011) A novel alpha-d-galactosynthase from *Thermotoga maritima* converts beta-d-galactopyranosyl azide to alpha-galacto-oligosaccharides. Glycobiology 21:448–456. https://doi.org/10.1093/glycob/cwq177

Cobucci-Ponzano B, Perugino G, Strazzulli A, Rossi M, Moracci M (2012) Thermophilic glycosynthases for oligosaccharides synthesis. Methods Enzymol 510:273–300. https://doi.org/10.1016/B978-0-12-415931-0.00015-X

Cobucci-Ponzano B, Strazzulli A, Iacono R, Masturzo G, Giglio R, Rossi M, Moracci M (2015) Novel thermophilic hemicellulases for the conversion of lignocellulose for second generation biorefineries. Enzyme Microb Tech doi:10.1016/j.enzmictec.2015.06.014

De Santi C, Tedesco P, Ambrosino L, Altermark B, Willassen NP, de Pascale D (2014) A new alkaliphilic cold-active esterase from the psychrophilic marine bacterium Rhodococcus sp.: functional and structural studies and biotechnological potential. Appl Biochem Biotechnol. 172:3054–3068. https://doi.org/10.1007/s12010-013-0713-1

Elleuche S, Schäfers C, Blank S, Schröder C, Antranikian G (2015) Exploration of extremophiles for high temperature biotechnological processes. Curr Opin Microbiol 25:113–119. https://doi.org/10.1016/j.mib.2015.05.011

Huang J, Zhang Y, Hu Y (2016) Functional Characterization of a Marine Bacillus Esterase and its Utilization in the Stereo-Selective Production of D-Methyl Lactate. Appl Biochem Biotechnol 180:1467–1481. https://doi.org/10.1007/s12010-016-2180-y

Iacono R, Cobucci-Ponzano B, Strazzulli A, Giglio R, Maurelli L, Moracci M (2016) (Hyper)thermophilic biocatalysts for second generation biorefineries. Chem Today 34:4

Iacono R, Strazzulli A, Maurelli L, Curci N, Casillo A, Corsaro MM, Moracci M, Cobucci-Ponzano B (2019) GlcNAc De-N-Acetylase from the Hyperthermophilic Archaeon *Sulfolobus solfataricus*. Applied and Environmental Microbiology 85 doi:10.1128/AEM.01879–18

Jaeger KE, Eggert T (2002) Lipases for biotechnology. Curr Opin Biotechnol 13:390–397

Jain I, Kumar V, Satyanarayana T (2014) Applicability of recombinant beta-xylosidase from the extremely thermophilic bacterium *Geobacillus thermodenitrificans* in synthesizing alkylxylosides. Bioresour Technol 170:462–469. https://doi.org/10.1016/j.biortech.2014.07.113

Jayanath G, Mohandas SP, Kachiprath B, Solomon S, Sajeevan TP, Bright Singh IS, Philip R (2018) A novel solvent tolerant esterase of GDSGG motif subfamily from solar saltern through metagenomic approach: Recombinant expression and characterization. Int J Biol Macromol 119:393–401. https://doi.org/10.1016/j.ijbiomac.2018.06.057

Kanamori Y, Watanabe M, Kawauchi K, Chen Y-G, Yanagishita H, Hirata H (2005) Kinetic Resolution of Enantiomers in Racemic and Enantiomerically Enriched 2-Alkanols by *Pseudomonas cepacia* Lipase Catalyzed Transesterification with Isopropenyl Acetate in Organic Solvent vol 54. doi:10.5650/jos.54.21,

Khan A, Singh P, Srivastava A (2018) Synthesis, nature and utility of universal iron chelator—Siderophore: a review. Microbiol Res 212–213:103–111. https://doi.org/10.1016/j.micres.2017.10.012

Kim J, Kim S, Yoon S, Hong E, Ryu Y (2015) Improved enantioselectivity of thermostable esterase from *Archaeoglobus fulgidus* toward (S)-ketoprofen ethyl ester by directed evolution and characterization of mutant esterases. Appl Microbiol Biotechnol 99:6293–6301. https://doi.org/10.1007/s00253-015-6422-7

Kumagai PS, Gutierrez RF, Lopes JLS, Martins JM, Jameson D, Castro AM, Martins LF, DeMarco R, Bossolan NRS, Wallace BA, Araujo APU (2018) Characterization of esterase activity from an *Acetomicrobium hydrogeniformans* enzyme with high structural stability in extreme conditions. Extremophiles: Life Under Extreme Conditions 22:781–793. https://doi.org/10.1007/s00792-018-1038-3

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis Version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054

Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) Procheck—a Program to Check the Stereochemical Quality of Protein Structures. J Appl Crystallogr 26:283–291. https://doi.org/10.1107/S0021889892009944

Levisson M, van der Oost J, Kengen SWM (2007) Characterization and structural modeling of a new type of thermostable esterase from *Thermotoga maritima*. Febs J 274:2832–2842. https://doi.org/10.1111/j.1742-4658.2007.05817.x

Levisson M, van der Oost J, Kengen SWM (2009) Carboxylic ester hydrolases from hyperthermophiles. Extremophiles: Life Under Extreme Conditions 13:567–581. https://doi.org/10.1007/s00792-009-0260-4

Li WQ, Shi H, Ding HH, Wang LL, Zhang Y, Li X, Wang F (2018) Characterization of two novel thermostable esterases from *Thermoanaerobacterium thermosaccharolyticum*. Protein Expres Purif 152:64–70. https://doi.org/10.1016/j.pep.2018.04.010

Li X, Yu H (2015) Characterization of an organic solvent-tolerant lipase from *Haloarcula sp*. G41 and its application for biodiesel production. Folia Microbiol 59(6):455–63. doi: 10.1007/s12223–014–0320–8

Lin H, Fischbach MA, Liu DR, Walsh CT (2005) In vitro characterization of salmochelin and enterobactin trilactone hydrolases IroD, IroE, and Fes. J Am Chem Soc 127:11075–11084. https://doi.org/10.1021/ja0522027

López-López O, Cerdán ME, González Siso MI (2014) New extremophilic lipases and esterases from metagenomics. Curr Protein Pept Sc 15:445–455

Manachini PL, Mora D, Nicastro G, Parini C, Stackebrandt E, Pukall R, Fortina MG (2000) *Bacillus thermodenitrificans* sp. nov., nom. rev. International journal of systematic and evolutionary microbiology 50 Pt 3:1331–1337 doi:10.1099/00207713–50–3–1331

Marcolongo L, La Cara F, Morana A, Di Salle A, Del Monaco G, Paixao SM, Alves L, Ionata E (2015) Properties of an alkali-thermo stable xylanase from *Geobacillus thermodenitrificans* A333 and applicability in xylooligosaccharides generation. World J Microbiol Biotechnol 31:633–648. https://doi.org/10.1007/s11274-015-1818-1

Nazina TN, Tourova TP, Poltaraus AB, Novikova EV, Grigoryan AA, Ivanova AE, Lysenko AM, Petrunyaka VV, Osipov GA, Belyaev SS, Ivanov MV (2001) Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, G. th. Int J Syst Evolut Microbiol 51:433–446. https://doi.org/10.1099/00207713-51-2-433

Ollinger J, Song KB, Antelmann H, Hecker M, Helmann JD (2006) Role of the Fur regulon in iron transport in *Bacillus subtilis*. J Bacteriol 188:3664–3673. https://doi.org/10.1128/JB.188.10.3664-3673.2006

Pereira MR, Maester TC, Mercaldi GF, de Macedo Lemos EG, Hyvonen M, Balan A (2017) From a metagenomic source to a high-resolution structure of a novel alkaline esterase. Appl Microbiol Biotechnol 101:4935–4949. https://doi.org/10.1007/s00253-017-8226-4

Perraud Q, Moynie L, Gasser V, Munier M, Godet J, Hoegy F, Mely Y, Mislin GLA, Naismith JH, Schalk IJ (2018) A Key Role for the Periplasmic PfeE Esterase in Iron Acquisition via the Siderophore Enterobactin in *Pseudomonas aeruginosa*. Acs Chem Biol 13:2603–2614. https://doi.org/10.1021/acschembio.8b00543

Ramnath L, Sithole B, Govinden R (2017) Classification of lipolytic enzymes and their biotechnological applications in the pulping industry. Can J Microbiol 63:179–192. https://doi.org/10.1139/cjm-2016-0447

Ranjan R, Yadav MK, Suneja G, Sharma R (2018) Discovery of a diverse set of esterases from hot spring microbial mat and sea sediment metagenomes. Int J Biol Macromol 119:572–581. https://doi.org/10.1016/j.ijbiomac.2018.07.170

Samoylova YV, Sorokina KN, Romanenko MV, Parmon VN (2018) Cloning, expression and characterization of the esterase estUT1 from *Ureibacillus thermosphaericus* which belongs to a new lipase family XVIII. Extremophiles : life under extreme conditions 22:271–285. https://doi.org/10.1007/s00792-018-0996-9

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75

Sirec T, Strazzulli A, Isticato R, De Felice M, Moracci M, Ricca E (2012) Adsorption of β-galactosidase of *Alicyclobacillus acidocaldarius* on wild type and mutants spores of Bacillus subtilis. Microb Cell Fact 11:100. https://doi.org/10.1186/1475-2859-11-100

Strazzulli A, Cobucci-Ponzano B, Carillo S, Bedini E, Corsaro MM, Pocsfalvi G, Withers SG, Rossi M, Moracci M (2017) Introducing transgalactosylation activity into a family 42 beta-galactosidase. Glycobiology 27:425–437. https://doi.org/10.1093/glycob/cwx013

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

Strazzulli A, Iacono R, Giglio R, Moracci M, Cobucci-Ponzano B (2017c) Metagenomics of hyperthermophilic environments: Biodiversity and biotechnology. In: Microbial Ecology of Extreme Environments. pp 103–135. doi:10.1007/978–3–319–51686–8_5

Wang L, Tang Y, Wang S, Liu RL, Liu MZ, Zhang Y, Liang FL, Feng L (2006) Isolation and characterization of a novel thermophilic *Bacillus* strain degrading long-chain n-alkanes. Extremophiles : life under extreme conditions 10:347–356. https://doi.org/10.1007/s00792-006-0505-4

Yang Z, Zhang Y, Shen T, Xie Y, Mao Y, Ji C (2013) Cloning, expression and biochemical characterization of a novel, moderately thermostable GDSL family esterase from *Geobacillus thermodenitrificans* T2. J Biosci Bioeng 115:133–137. https://doi.org/10.1016/j.jbiosc.2012.08.016

Yang X, Wu L, Xu Y, Ke C, Hu F, Xiao X, Huang J (2018) Identification and characterization of a novel alkalistable and salt-tolerant esterase from the deep-sea hydrothermal vent of the East Pacific Rise. Microbiologyopen 7:e00601. https://doi.org/10.1002/mbo3.601

Yu N, Yang JC, Yin GT, Li RS, Zou WT, He C (2018) Identification and characterization of a novel esterase from *Thauera sp*. Biotechnol Appl Bioc 65:748–755. https://doi.org/10.1002/bab.1659

Zhang J, Zhao M, Yu D, Yin J, Zhang H, Huang X (2017) Biochemical characterization of an enantioselective esterase from *Brevundimonas* sp. LY-2. Microb Cell Fact 16:112 doi:10.1186/s12934-017-0727-4

Zhu Y, Li J, Cai H, Ni H, Xiao A, Hou L (2013) Characterization of a new and thermostable esterase from a metagenomic library. Microbiol Res 168:589–597. https://doi.org/10.1016/j.micres.2013.04.004

Zhu Y, Zheng W, Ni H, Liu H, Xiao A, Cai H (2015) Molecular cloning and characterization of a new and highly thermostable esterase from *Geobacillus* sp. JM6. J Basic Microbiol 55:1219–1231. https://doi.org/10.1002/jobm.201500081

3.3. Conclusion

Characterized enterochelin/enterobactin esterases are still poorly biochemically characterized, leaving their biotechnological potential unexploited. In the study reported, the enterochelin/enterobactin group has been classified as a lipolytic enzyme family for the first time. The enzymatic characterization of EstGtA3 showed a behaviour like the canonical esterases with interesting features suitable for resolving racemic mixtures and synthesizing optically pure compounds. This study demonstrates that even enzymes that work in a peculiar metabolic pathway, which was never considered for biotechnological applications, could have suitable features for these purposes. In addition, considering the vast number of sequences annotated in the online databases, the research of exciting enzymes for industrial processes by genome mining turns out to be a valid alternative for discovering new biocatalysts.

CHAPTER 4

4. Xyloglucan-active enzymes from *S. solfataricus*

4.1. Overview

Carbohydrates active enzymes play an important role in the shift towards a bio-sustainable circular economy or as described in the thesis Introduction, a circular bio-economy. GHs are involved in the conversion of lignocellulosic biomass to produce sugars that can be fermented in bioethanol and high value-added products. The process, as previously described, involves the saccharification of cellulose and hemicellulose. The major hemicelluloses including xylan, mannan and xyloglucan, and, after cellulose, they are the second most abundant polymer in plants, one-quarter of the global dry plant biomass, and represent one of the most heterogeneous polymers on Earth [143]. Indeed, their classification is based on the predominant monosaccharides of the backbone formed by an equatorial chain of monosaccharidic units, that are highly branched by sugars linked to each other with both equatorial and axial bonds. Their complex structure makes them recalcitrant to spontaneous hydrolysis and several glycosidase activities are necessary for their degradation. To achieve the complete saccharification of both cellulose and hemicellulose it is necessary to develop an efficient process of lignocellulose feedstock biotransformation [144]. In this frame, this chapter describes the characterization of the mechanism of action on xyloglucan oligosaccharides of three GHs from the hyperthermophilic archaeon *Saccharolobus solfataricus*. In this chapter I will report the entire work in which my contribution has been predominant**.**

4.1.1. Saccharolobus solfataricus

S. solfataricus is an aerobic thermoacidophilic archaeon, representing a recognized model system for biochemistry and molecular biology studies in the Crenaerchota phylum. Previously belonging to the genus Sulfolobus, after the description of *Saccharolobus caldissimus*, in 2018 it was reclassified to the genus of Saccharolobus [145]. The hyperthermophilic archaeon optimally thrives at 80 °C (temperature range of 60-92 °C) and pH range 2.0-4.0. It was first isolated from the Pisciarelli Solfatara by the German microbiologists Karl Setter and Wolfram Zillig in the 1980 [146]. However, the microorganism populates all hot acid environment around the world. *S. solfataricus* can use as carbon and energy sources for many organic compounds, such as carbohydrates, amino acids, or peptides it is able to metabolize sulphur and can be cultured in laboratory relatively easily [147]. The genome of *S. solfataricus* contains many sequences encoding for GHs. Indeed, it was a source of a variety of (hemi)cellulose-degrading enzymes such

as xylanase, glucoamylases, cellulases, α-amylase, β-glucosidase, α-xylosidase and α-fucosidase [148-154].

4.1.2. Xyloglucan

Xyloglucan (XG) hemicellulose is widely distributed in plants. It is the principal hemicellulose component in the dicot primary cell wall (20%–30% of the total cell wall) and one of the most abundant storage polysaccharides in seeds (>40% in weight in some species). This polymer is composed of a β-(1,4)-glucan backbone, with α-(1,6)-*D*-xylose groups linked to about 75% of the glucosyl residues. A degree of complexity is added in some plant species by the presence of other differently linked monosaccharides such as α-*L*-Fuc and β-*D*-Gal, among the others. This hemicellulose displays a great heterogeneity, in fact its composition depends on the species and the degree of plant ripeness [155,156]. XG is an interesting source of monosaccharides in biorefineries, but it has also interesting chemico-physical properties. XG is a hydrocolloid in solution showing low viscosity and low molecular weight compared to many other polysaccharides, forming a gel at high sugar concentration and low water activity condition. For its properties, XG and its chemically/enzymatically modified oligosaccharidic derivatives have a wide range of applications [157]. Among the lignocellulose bioconversion it has been demonstrated that the presence of endo-xyloglucanases improve the yield of saccharification [158]. Nevertheless, the development of enzymatic cocktails degrading XG is still in its infancy [159].

International Journal of
Molecular Sciences

MDPI

Article

Xyloglucan Oligosaccharides Hydrolysis by Exo-Acting Glycoside Hydrolases from Hyperthermophilic Microorganism *Saccharolobus solfataricus*

Nicola Curci [1,2], Andrea Strazzulli [1,3], Roberta Iacono [1], Federica De Lise [2], Luisa Maurelli [2], Mauro Di Fenza [2], Beatrice Cobucci-Ponzano [2,*] and Marco Moracci [1,3]

[1] Department of Biology, University of Naples "Federico II", Complesso Universitario di Monte S. Angelo, 80126 Naples, Italy; nicola.curci@unina.it (N.C.); andrea.strazzulli@unina.it (A.S.); roberta.iacono@unina.it (R.I.); marco.moracci@unina.it (M.M.)
[2] Institute of Biosciences and BioResources—National Research Council of Italy, 80131 Naples, Italy; federica.delise@ibbr.cnr.it (F.D.L.); luisa.maurelli@ibbr.cnr.it (L.M.); mauro.difenza@ibbr.cnr.it (M.D.F.); beatrice.cobucciponzano@ibbr.cnr.it (B.C.-P)
[3] Task Force on Microbiome Studies, University of Naples Federico II, 80134 Naples, Italy
* Correspondence: beatrice.cobucciponzano@ibbr.cnr.it

Abstract: In the field of biocatalysis and the development of a bio-based economy, hemicellulases have attracted great interest for various applications in industrial processes. However, the study of the catalytic activity of the lignocellulose-degrading enzymes needs to be improved to achieve the efficient hydrolysis of plant biomasses. In this framework, hemicellulases from hyperthermophilic archaea show interesting features as biocatalysts and provide many advantages in industrial applications thanks to their stability in the harsh conditions encountered during the pretreatment process. However, the hemicellulases from archaea are less studied compared to their bacterial counterpart, and the activity of most of them has been barely tested on natural substrates. Here, we investigated the hydrolysis of xyloglucan oligosaccharides from two different plants by using, both synergistically and individually, three glycoside hydrolases from *Saccharolobus solfataricus*: a GH1 β-gluco-/β-galactosidase, a α-fucosidase belonging to GH29, and a α-xylosidase from GH31. The results showed that the three enzymes were able to release monosaccharides from xyloglucan oligosaccharides after incubation at 65 °C. The concerted actions of β-gluco-/β-galactosidase and the α-xylosidase on both xyloglucan oligosaccharides have been observed, while the α-fucosidase was capable of releasing all α-linked fucose units from xyloglucan from apple pomace, representing the first GH29 enzyme belonging to subfamily A that is active on xyloglucan.

Keywords: glycoside hydrolases; *Saccharolobus solfataricus*; xyloglucan; polysaccharide degradation; archaea

Citation: Curci, N.; Strazzulli, A.; Iacono, R.; De Lise, F.; Maurelli, L.; Di Fenza, M.; Cobucci-Ponzano, B.; Moracci, M. Xyloglucan Oligosaccharides Hydrolysis by Exo-Acting Glycoside Hydrolases from Hyperthermophilic Microorganism *Saccharolobus solfataricus*. *Int. J. Mol. Sci.* **2021**, *22*, 3325. https://doi.org/10.3390/ijms22073325

Academic Editor: Kunio Takeyasu

Received: 2 March 2021
Accepted: 22 March 2021
Published: 24 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

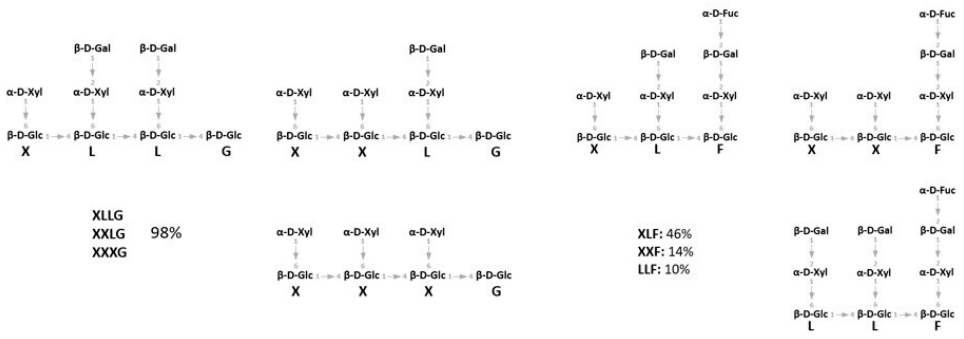
1. Introduction

Hemicelluloses include heterogeneous polymers characterized by a linear backbone with an equatorial configuration highly substituted by short oligosaccharides, monosaccharides, and organic acids. Hemicellulose is the second most abundant polysaccharide in the plant cell wall, representing about 20–35% of lignocellulose biomasses. Its complete degradation requires the simultaneous action of various enzymatic activities due to the heterogeneous composition in monosaccharides, leading to high variability in its structure [1].

Hemicellulose-degrading enzymes belong to the classes of glycosyl hydrolases (GH) and carbohydrate esterase (CE) which are classified in the carbohydrate active enzymes (CAZy, www.cazy.org (accessed on 2 March 2021)) database in functional families based

on amino acidic sequences and structure-related similarities [2]. These enzymes have applications in many biotechnological processes, such as biorefineries for the conversion of lignocellulose biomasses in biofuels and bioplastic precursors [3]. Many industrial biomass-degrading processes are performed at high temperatures and extreme pH, to improve the solubility and the availability of organic compounds. In this regard, hemicellulases from (hyper)thermophiles, compared to their mesophilic counterparts, provide many advantages in industrial processes thanks to their high thermal operational stability and tolerance to solvents [4]. In particular, hyperstable carbohydrate active enzymes (cazymes) from (hyper)thermophilic archaea show maximum activity at higher temperature ranges compared to mesophilic cazymes. However, archaeal GHs are poorly represented in CAZy and less studied compared to their bacterial counterparts [5]. Moreover, although most of the GHs annotated in the CAZy database have been characterized by using synthetic substrates, in the framework of biomass saccharification, it is important to test their activity on natural substrates and, more importantly, to investigate their synergistic action on this type of substrate. Cocktails of cazymes need to be improved to achieve a more efficient biotransformation of lignocellulose biomasses [6,7]. Polysaccharides from lignocellulose biomasses can be very heterogeneous in structure and chemical composition, because they can originate from different parts of plants with different degrees of maturation and from a variable number of species.

Xyloglucan (XG) hemicellulose is present in all land plants. It is among the major polysaccharides of the Type I primary cell wall of all dicots and a great number of monocots [8]. It is composed of a 1,4-β-glucan backbone frequently substituted at O-6 by α-*D*-xylose residue. Xylose is typically the first residue of XG sidechains, which are substituted by up to four glycoside residues, including β-*D*-galactose, which is itself substituted by α-*L*-fucose in some plant species. The sidechains of XG are described by a single letter code, in which G represents the unsubstituted glucose in the backbone, whereas X, L and F represent the glucose residues substituted with α-*D*-xylosyl, β-*D*-galactosyl-(1,2)-α-*D*-xyloside, and α-*L*-fucosyl-(1,2)-β-*D*-galactosyl-(1,2)-α-*D*-xyloside sidechains, respectively (Figure 1) [9,10].



(**a**) (**b**)

Figure 1. (**a**) Schematic structure of xyloglucan (XG) oligosaccharides (XGOs) from tamarind seeds consisting of a mix of XGOs (XLLG, XXLG, XXXG). (**b**) Schematic structure of XGOs from apple pomace consisting of a mix of XGOs with structure XLF, XXF, LLF (see text for the nomenclature).

To date, nineteen types of XG sidechain are described, and a large number of enzymes involved in the degradation of XG have been identified and characterized in fungi [8,11] and in Gram-negative and -positive bacteria, such as human intestinal *Bacteroides*, saprophytic *Cellvibrio japonicus* and anaerobic *Ruminoclostridium cellulolyticum* [12–14]. In striking contrast, the degradation of XG by archaeal enzymes has been poorly studied. The complete saccharification of XG to component monosaccharides requires the

concomitant action of a backbone-cleaving endo-xyloglucanase and exo-glycosidases on the non-reducing ends of XG oligosaccharides (XGOs) [15].

The chemoheterotroph thermoacidophile archaeon *Saccharolobus solfataricus* (previously *Sulfolobus solfataricus*) has also been reclassified for its ability to grow on a variety of carbon sources including sugars, such as polysaccharides (cellulose, starch, dextrin), disaccharides (maltose, sucrose, lactose), hexoses (e.g., D-galactose, D-glucose, D-mannose and L-fucose), and pentoses (e.g., D-xylose and L-arabinose) monosaccharides [16–18]. This (hyper)thermophilic microorganism, one of the best-studied in the *Crenarchaeota* phylum, thrives in terrestrial volcanic hot springs and has optimal growth at 80 °C and pH 2–4. In these extreme habitats, lignocellulosic biomass from the nearby vegetation can experience chemical transformation in the hot and acid conditions, providing carbon sources to the microbial community of extremophiles. Consistently, several sequences encoding for GH and mono-, di- and oligo-transporters for sugars uptake are present in the genome of *S. solfataricus* [19,20]. Many of the characterized GHs from *S. solfataricus* showed potential hemicellulolytic activities (GH1, GH2, GH3, GH5, GH12, GH29, GH31, GH36, GH38, GH116), and even some genetic clusters could suggest the ability to hydrolyze hemicelluloses (e.g., *S. solfataricus* P2 Open reading frames (SSO), SSO1353 and SSO1354 encoding for enzymes with β-glucosidase/β-xylosidase and β-glucanase/β-xylanase activity, respectively) [21–32]. However, to date, patterns involved in hemicellulose hydrolysis have never been described.

In this work, we tested the activity of three exo-acting GHs from *S. solfataricus* on XGOs from tamarind seeds and apple pomace that we named XGO1 and XGO2, respectively (Figure 1). These enzymes were already biochemically characterized and showed potential activity for the degradation of XGOs: GH1 (LacS) displayed β-glucosidase and β-galactosidase activities, while GH31 (XylS) and GH29 (SsαFuc) showed α-xylosidase and α-fucosidase activity, respectively. LacS and XylS already showed the capability to hydrolyze in a concerted manner short and low-branched XGOs from tamarind seed, while SsαFuc was able to hydrolyze short fucosylated oligosaccharides [21,23,25]. We show here the mechanism of action of the three enzymes on XGOs that allowed us to propose their possible function in vivo.

2. Results

2.1. Protein Analysis of Recombinant GHs

Recombinant LacS, XylS and SsαFuc were purified with a final yield of 60 mg L^{-1}, 1 mg L^{-1} and 1.5 mg L^{-1}, respectively. The specific activity of the enzymes was evaluated by using aryl-glycosides as substrates at 65 °C in sodium acetate pH 5.5. LacS had a specific activity of 102 U mg^{-1} and 29 U mg^{-1} on 5 mM p-nitrophenyl-βGlc (pNP-βGlc) and 5 mM o-nitrophenyl-βGal (oNP-βGal), respectively (in the following text we will refer to the LacS units on the pNP-βGlc). XylS had a specific activity of 2 U mg^{-1} on 32 mM pNP-αXyloside, and SsαFuc had a specific activity of 22 U mg^{-1} on 2 mM pNP-αFucoside.

2.2. Hydrolysis of XGO from Tamarind Seeds

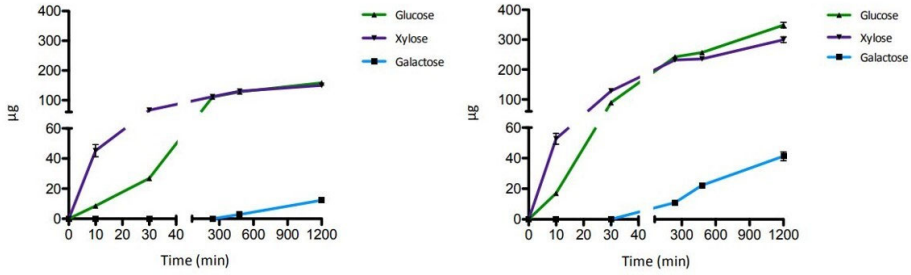
For testing, we have used XGOs from tamarind seeds (XGO1) that were composed of XXLG, XXXG, and XLLG according to XG nomenclature (Figure 1). The simultaneous and singular activity of LacS, at two concentrations, and XylS on XGO1 have been investigated after 20 h of incubation at 65 °C in sodium acetate buffer, pH 5.5 (Table 1). Moreover, the synergistic activity of the two enzymes was also evaluated after incubations of 10 and 30 min and 4, 8 and 20 h, in the same conditions (Figure 2). Firstly, the activity on XGO1 was evaluated by using 2.2 U for LacS and 0.1 U for XylS. Their simultaneous activity showed the release of galactose, glucose and xylose (12.3 μg, 158.3 μg and 149.7 μg, respectively) (Table 1; Figure S1). XylS alone catalyzed the release of 89.1 μg of xylose, while, in combination with LacS, the amount of xylose increased up to 149.7 μg. LacS action did not lead to any product when acting alone on XGO1, and only the use in synergy with

XylS allowed the release of the glucose and galactose. Xylose and glucose were detectable after 10 min, while galactose became detectable only after 8 h of incubation (Figure 2a, Figure S2 and Table S1). Thereafter, in order to understand if increased amounts of LacS could improve the efficiency of hydrolysis of XGO1, we used about eightfold more LacS (18 U), keeping constant the XylS units. Even in this case, LacS alone did not produce detectable amounts of galactose and glucose. However, the amounts of all the monosaccharides released (Table 1, Figure S3) were significantly increased, by 3.4, 2.2, and 2 times, for galactose, glucose and xylose, respectively. The time course of the synergistic action of the two GHs (Figure 2b, Figure S4 and Table S2) showed that at all the incubation times, the amounts of monosaccharides detected were higher if compared to those obtained with 2.2 units of LacS (Figure 2a), and that galactose could be detected after 4 h of incubation. Moreover, release of the monosaccharides increased over time until 20 h. Although this process was not complete compared to the acid hydrolysis, probably due to product inhibition of one or both enzymes, it led to a conversion of 10%, 41% and 37% for galactose, glucose and xylose, respectively (Table 1). This is a remarkable bioconversion efficiency, and demonstrates the good processivity and the extreme thermal operation stability of LacS and XylS on this substrate.

Table 1. Enzymatic hydrolysis of XGO1 after 20 h by using LacS and XylS.

Enzymes	Galactose (μg)	Glucose (μg)	Xylose (μg)
TFA hydrolysis [1]	365 ± 16	835 ± 26	803 ± 9
LacS [2] + XylS	12.3 ± 1.1	158.3 ± 1.3	149.7 ± 1
LacS [3] + XylS	41.3 ± 2.9	348.9 ± 9.3	299.9 ± 9.6
LacS[1]	ND	ND	ND
LacS[2]	ND	ND	ND
XylS	ND	ND	89.1 ± 9

Trifluoracetic acid (TFA) hydrolysis [1]: 2 h at 100 °C in 2 M of TFA; LacS [2]: 2.2 units; LacS [3]: 18 units.



(**a**) (**b**)

Figure 2. Time course of the enzymatic hydrolysis of XGO1. (**a**) LacS 2.2 U and XylS 0.1 U. (**b**) LacS 18 U and XylS 0.1 U. Error bars reported represent standard deviations.

2.3. Hydrolysis of XGO from Apple Pomace

The XGOs purified from apple pomace (XGO2) were composed of XLF, XXF, and LLF according to the XG nomenclature (Figure 1b) and showed a 70% degree of purity; the major contaminants were represented by arabinan and chitin as reported in the manufacturing notes. The enzymatic degradation of XGO2 was analyzed after 20 h of incubation at 65 °C in sodium acetate buffer (pH 5.5) by combining 0.3 U, 0.1 U and 2.2 U of SsαFuc, XylS and LacS, respectively, together, in pairs, and singly. Table 2 reports the quantitative analysis of the reaction products of the enzymatic and chemical conversion. The quantifications of the single monosaccharides by amperometric detector (HPAEC-

PAD) showed that the three enzymes could release fucose, galactose, glucose, and xylose from this substrate (Figure S5). When the three enzymes acted simultaneously, xylose and fucose were the predominant monosaccharides released, while galactose and glucose were produced in lower quantities. When SsαFuc acted alone or in combination with the other enzymes the amount of fucose released did not vary significantly, suggesting that the enzyme was able to remove the fucosidic residues from XGO2 with no need of previous or concomitant action by the other two enzymes. Interestingly, the amount of fucose released enzymatically was 90% of that obtained by complete chemical hydrolysis of XGO2, indicating that SsαFuc hydrolyzed the substrate with high efficiency. Instead, xylose, glucose, and galactose were released by the three enzymes as 37%, 10% and 3%, respectively, of the monosaccharides identified after chemical treatment. The simultaneous action of XylS and LacS resulted in higher amounts of glucose and xylose compared to those obtained when the two enzymes acted singly or when each of them were separately combined with SsαFuc, and similar to when the three enzymes acted together.

Table 2. Enzymatic hydrolysis of XGO2 after 20 h by using LacS, XylS and SsαFuc.

Enzymes	Fucose (μg)	Galactose (μg)	Glucose (μg)	Xylose (μg)
TFA hydrolysis [1]	83 ± 2	220 ± 18	346 ± 25	306 ± 11
LacS + XylS + SsαFuc	75.5 ± 5	6.6 ± 1	34.7 ± 5.8	115 ± 13
LacS + XylS	ND	3.15 ± 0.1	35 ± 0.5	118 ± 3.9
XylS +SsαFuc	80.1 ± 14	ND	ND	72.8 ± 9
LacS + SsαFuc	79.5 ± 9	4.7 ± 0.3	24.2 ± 4	ND
SsαFuc	73.3 ± 11	ND	ND	ND
LacS	ND	1.5	13.6	ND
XylS	ND	ND	ND	73.5 ± 11

TFA hydrolysis [1]: 2 h at 100 °C in 2 M of TFA.

The time course of the synergistic action of the three GHs was evaluated after incubations of 10 and 30 min and 4, 8 and 20 h. As shown in Figure 3, the amount of galactose, glucose, and xylose released by enzymatic hydrolysis increased until the 20 h mark, while the fucose release reached a plateau after 4 h (Table S3 and Figure S6). A small amount of galactose (0.5 μg) was detectable only after 30 min of incubation. These results confirmed the high efficiency of SsαFuc and the thermal operational stability of the three enzymes at 65 °C on the substrate tested.

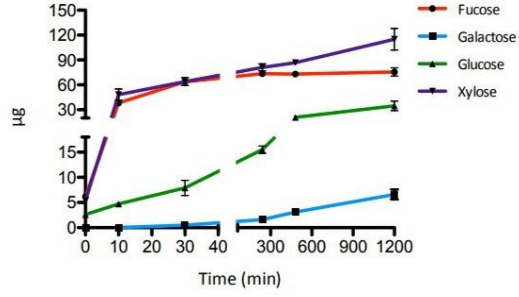


Figure 3. Time course of the enzymatic hydrolysis of XGO2. Error bars reported represent standard deviations.

3. Discussion

In this study, we monitored the release of monosaccharides from XGOs, using different combinations of three GHs from *S. solfataricus* in order to investigate the mode of action of these enzymes on xyloglucan oligosaccharides. The substrates used differed in composition and degree of purity. XGO1, from tamarind seeds which lack α-fucosyl residues, was used to better understand the mechanism of action of XylS and LacS on the X and L components. Instead, XGO2 purified from apple pomace, in which 70% of sidechains showed α-linked fucose residues, was used to investigate the action of SsαFuc on F components and the simultaneous hydrolysis of the α-fucosidase with LacS and XylS.

The analysis of the enzymatic hydrolysis of XGO1 confirmed the concerted action of the LacS and XylS exo-glycosidases, catalyzing the processive hydrolysis of α1,6-linked xylose, β1,4-linked glucose and β1,2-linked galactose from the non-reducing end of XGOs. In particular, XylS removed only the first xyloside at the non-reducing end of XGOs, as previously shown on other xylosylated oligosaccharides [21,33] and as reported for other GH31 α-xylosidases on xyloglucan [34]. Therefore, XylS requires the removal of glucose and galactose residues at the non-reducing end to access to the following xyloside residues (Figure 1). LacS, like other GH1 enzymes [34], has broad substrate specificity, promoting the hydrolysis of 1,4- and 1,2-β-*D*-glycosides [23,35,36]. Moreover, an in silico docking study by Kumar and co-workers suggested that a LacS active site could well accommodate short-branched XGOs (GLG) with β-galactose at non-reducing termini in the subsite -1, and the interaction with XGOs could be stronger and selective compared to that of the XG-active β-galactosidase of *C. japonicus* [35]. Indeed, the simultaneous activity of XylS and LacS on both XGO1 and XGO2 produced more xylose compared to the activity of the α-xylosidase alone. In principal, LacS could act on both the β-*D*-linked galactose and glucose residues of XGOs substrates. The substrates used in this work showed galactose at the non-reducing end of XGO branches. Glucose, on the other hand, being decorated by xylose residues, was not accessible to the enzyme (Figure 1). Thus, LacS could operate on galactose residues, while access to bound glucosides would occur only after the previous hydrolysis of 1,6-α-*D*-xylosides promoted by XylS. As expected, LacS alone on XGO1 could not catalyze the release of glucose. However, galactose was not observed either, even by using increasing LacS enzymatic units. This suggests that in order to access to the β1,2-linked galactose, LacS needed partial hydrolysis of the substrate operated by the synergistic and processive action of XylS and LacS itself (Figure 4).

The time courses in Figure 2 show that galactose is released only after 4 h and 8 h of enzymatic incubation, by using 2.2 or 18 units of LacS, respectively. In particular, by increasing the enzymatic units of LacS, even keeping the enzymatic units of XylS constant, the released amounts of glucose and galactose, as well as that of xylose, increased significantly, with an earlier release of galactose resulting in more efficient hydrolysis of XGO1. Higher amounts of LacS offered a greater availability of hydrolyzable substrate to XylS, confirming also, in this case, the synergistic action of the two enzymes.

The GH29 α-fucosidase from *S. solfataricus* released 96% of all α1,2-linked fucose available in XGO2 according to the complete chemical hydrolysis used as control. This indicates that the enzyme was able to access and remove with high efficiency the fucosides decorating XGO2 with no need of prior substrate debranching, indicating the ability of SsαFuc to recognize complex fucosylated oligosaccharides. Most of the α-fucosidases are currently classified in the CAZy database in two major families: GH29 and GH95 [2]. The characterized enzymes of family 95, which also include α-galactosidases, have no activity on *p*NPα-Fuc but are active on α-1,2-linked fucose of XG [37,38]. Instead, according to the classification of Sakurama et al. (2012), the α-fucosidases of the GH29 family were grouped in two subfamilies: GH29A, which are active on *p*NPα-Fuc and several fucosylated oligosaccharides, and GH29B, which are specific for α-1,3- and α-1,4-linked fucosides, but not *p*NPα-Fuc [39]. The GH29B enzyme from *Fusarium graminearum* is active on fucosylated XGOs [40].

SsαFuc belongs to the subfamily GH29A, in which the enzymes display activity toward any α-linked fucose (α-1,2; α-1,3; α-1,4; α-1,6) [41,42]. However, activity on fucosylated XGOs has not been so far observed for GH29A enzymes. We have previously shown that SsαFuc hydrolyzed a short synthetic fucosylated aryl disaccharide α-*L*-Fucosyl(1-3)-α-*L*-Fucosyl-O-*p*NP [25]. Here, we demonstrated for the first time the specificity of SsαFuc for α1,2-linked fucosides on a natural substrate. SsαFuc is the only known characterized GH29A active on fucosylated XGOs. The hydrolytic activity of SsαFuc on XGO1 produced XGOs free of α-fucosyl residues that are a suitable substrate for LacS and XylS.

The data obtained allow us to propose a possible function in vivo of the three GHs from *S. solfataricus*. It is still unclear whether this archaeon could hydrolyze and use XG as a carbon source, as neither xyloglucanase nor XGOs transporters have ever been identified in *S. solfataricus*. To date, the characterized α-xylosidases of the GH31 family are always implicated in XG metabolism and, in the well-characterized XG utilization pathway of *Bacteroidetes* and *C. japonicus*, α-fucosidase activity is involved in XG degradation and utilization [12,14,15,34,37,43–47]. Moreover, XylS and LacS genes mapped in the same locus, in a 50 kb genome region also including the SsαFuc gene [21]. In light of the strict synergism between LacS and XylS, and the high efficiency of SsαFuc to release Fuc from XGOs, the results obtained suggest that the three enzymes could play a role in vivo in the degradation of xyloglucan as an energy source in *S. solfataricus*. Furthermore, from a biotechnological point of view, this study sheds light on the mode of action of the three enzymes on XGOs, laying the foundations for the possible development of a thermostable enzymatic cocktail to hydrolyze XG.

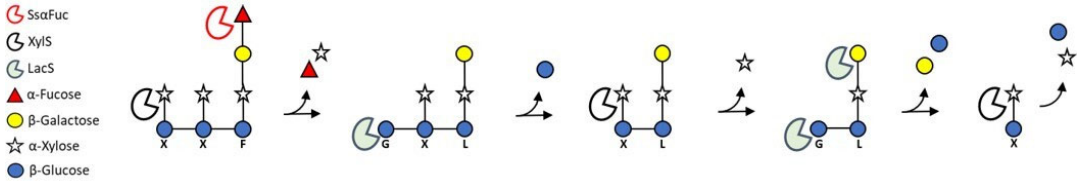


Figure 4. Proposed model of the reaction mechanism on XGOs by XylS, SsαFuc and LacS.

4. Materials and Methods

4.1. Substrates

Xyloglucan oligosaccharide from tamarind seeds (XGO1) with 98% of purity was purchased from Megazyme (Bray, Ireland). Xyloglucan oligosaccharide from apple pomace (XGO2) with 70% of purity was purchased from Elicityl OligoTech (Crolles, France). The major contaminants are represented by arabinan and chitin.

4.2. Enzyme Expression and Purification

The genes encoding for XylS, LacS and SsαFuc were expressed and the proteins purified as previously reported [21,23,25] with the following modifications. The gene encoding for XylS was carried by the plasmid pT7SCII-XylS under the control of an isopropyl -β-D-1-thio galactopyranoside (IPTG)-inducible T7 RNA polymerase promoter and was expressed in *E. coli* BL21 (DE3). The transformed cells were grown in 2 L of Super Broth (SB) supplemented with ampicillin 50 μg mL^{-1}. The gene expression was induced with 1 mM of IPTG (Merck, Darmstadt, Germany) at 0.5 OD_{600}. XylS was purified by three subsequent cell extract heating steps of 30 min at 55 °C, 65 °C, and 75 °C, followed by anion exchange chromatography. The gene encoding for LacS was carried by the plasmid pET29-LacS under the control of an IPTG-inducible T7 RNA polymerase promoter and was expressed in *E. coli* BL21(DE3). The transformed cells were grown in 2 L of SB supplemented with kanamycin 50 μg mL^{-1} (Merck, Darmstadt, Germany). The gene

expression was induced with 1 mM of IPTG at 0.5 OD_{600}. LacS was purified by three subsequent cell extract heating steps at 55 °C, 65 °C, 75 °C for 30 min and 85 °C for 10 min, followed by hydrophobic exchange chromatography. The gene encoding for SsαFuc was carried by pGex-SsαFuc plasmid under the control of an IPTG-inducible TAC promoter, and was expressed in *E. coli* RB791 as a fusion protein with Glutathione-S-transferases-tag (GST-tag). The transformed cells were grown in 2 L of SB supplemented with ampicillin 50 μg mL^{-1} (Merck, Darmstadt, Germany). The gene expression was induced with 1 mM of IPTG at 0.5 OD_{600}. The recombinant enzyme was purified by affinity chromatography on Glutathione Sepharose 4B resin (GE Healthcare, Chicago, IL, USA) and the removal of GST-tag was obtained by thrombin cleavage.

4.3. Protein Analysis of the Recombinant Enzymes

The protein concentration of purified enzymes was measured by the Bradford method using bovine serum albumin (BSA) as standard [48]. Specific activities of the enzymes were measured by using aryl substrates *p*NPα-Xyl, *p*NPβ-Glc, *o*NP-βGal and *p*NPα-Fuc (Carbosynth, St. Gallen, Switzerland) for XylS, LacS and SsαFuc, respectively, in sodium acetate pH 5.5, at 65 °C. The releasing of *p*NP or *o*NP was monitored continuously at 405 nm using a double beam spectrophotometer with a thermal control unit (Varian Cary 100, Agilent Technologies, Santa Clara, CA, USA). The background hydrolysis of the substrates was subtracted by using reference samples identical to the reaction mixtures without enzymes. One unit (U) of activity was defined as the amount of enzyme that released 1 μmol of *p*NP or *o*NP (molar extinction coefficient of 2.08 and 1.2 mM^{-1} cm^{-1}, respectively) per minute at the standard conditions.

4.4. XGO1 and XGO2 Enzymatic Hydrolysis

A solution of XGO2 (1.1 mg mL^{-1}) was incubated with LacS (2.2 U), XylS (0.1 U) and SsαFuc (0.3 U), individually or simultaneously, in 100 mM of sodium acetate buffer, pH 5.5. A solution of XGO1 1.1 mg/mL^{-1} was incubated with LacS 2.2 U or 18 U and XylS 0.1 U, individually or simultaneously, in 100 mM sodium acetate buffer pH 5.5 at 65 °C. "Simultaneous reaction" refers to the assays performed with two or three enzymes used simultaneously. The blank mixtures had identical compositions to the reaction mixtures, using the enzyme storage buffers instead of the enzymes. The assays were performed at 65 °C, and after 10 and 30 min and 4, 8 and 20 h of incubation were stopped through instantaneous freezing in dry ice. All reaction products were analyzed and quantified by high-performance anion-exchange chromatography with pulsed amperometric detector (HPAEC-PAD) analysis (Dionex ICS 3000, Waltham, MA, USA).

4.5. Chemical Hydrolysis

XGO1 and XGO2 were dissolved in 2 M of trifluoracetic acid (TFA) at the concentration of 1.1 mg/mL and incubated for 2 h at 100 °C [49]. The reaction solutions were all dried by rotary evaporation to remove TFA. Then, dry pellets were resuspended in distilled water at 1.1 mg mL^{-1} and the drying process was repeated until TFA was completely removed. The monosaccharides composition was determined by HPAEC-PAD.

4.6. HPAEC-PAD Analysis

The monosaccharide composition of oligosaccharides XGO1 and XGO2 hydrolyzed via enzymatic and TFA hydrolysis was analyzed by an HPAEC-PAD system equipped with Carbopac PA-100 (Dionex, Sunnyvale, CA, USA). For the assays on XGO1 as substrate, glucose, fucose, galactose and xylose (Merck, Darmstadt, Germany) were used as standards to calibrate the retention time and to build a calibration line at concentrations of 0.5 nmol 25 μL^{-1}, 1.0 nmol 25 μL^{-1}, and 1.5 nmol 25 μL^{-1}, with arabinose (Merck, Darmstadt, Germany) as internal standards at a concentration of 0.5 nmol 25 μL^{-1}. For the

assay on XGO2, glucose, xylose and galactose were used as standards to calibrate the retention time and to build a calibration line at concentrations of 1 nmol 25 μL^{-1}, 2 nmol 25 μL^{-1}, and 3 nmol 25 μL^{-1}, with fuc as internal standards at concentration of 2 nmol 25 μL^{-1}. In total, 25 µL of the samples, supplemented with 0.5 nmol arabinose or 2 nmol of fucose, were injected for the HPAEC-PAD analysis. The elution program consisted of an isocratic elution with NaOH 8 mM for 20 min.

Supplementary Materials: The following are available online at www.mdpi.com/1422-0067/22/7/3325/s1, Figure S1. HPAEC-PAD chromatograms of the enzymatic treatment of XGO1 by using LacS 2.2 U and XylS 0.1 U. Figure S2. HPAEC-PAD chromatograms of the time course of the enzymatic treatment of XGO1 by using LacS 2.2 U and XylS 0.1 U. Figure S3. HPAEC-PAD chromatograms of the enzymatic treatment of XGO1 by using LacS 18 U and XylS 0.1 U. Figure S4. HPAEC-PAD chromatograms of the time course of the enzymatic treatment of XGO1 by using LacS 18 U and XylS 0.1 U. Figure S5. HPAEC-PAD chromatograms of the enzymatic treatment of XGO2 by LacS 2.2 U, XylS 0.1 U and Ss α Fuc 0.3 U. Figure S6. HPAEC-PAD chromatograms of the time course of the enzymatic treatment of XGO2 by LacS 2.2 U, XylS 0.1 U and SsαFuc 0.3 U. Table S1. Monosaccharides composition from the time course of the enzymatic treatment of XGO1 by using LacS 2.2 U and XylS 0.1 U. Table S2. Monosaccharides composition from the time course of the enzymatic treatment of XGO1 by using LacS 18 U and XylS 0.1 U. Table S3. Monosaccharides composition from the time course of enzymatic treatment of XGO2 by LacS 2.2 U, XylS 0.1 U and SsαFuc 0.2 U.

Author Contributions: conceptualization, N.C., M.M., B.C.-P., A.S.; methodology N.C.; validation, M.M., A.S., B.C.-P.; data curation, N.C.; writing—original draft preparation N.C.; writing—review and editing, A.S., M.M., B.C.-P., R.., F.D.L., M.D.F., L.M.; funding acquisition, M.M. and B.C.-P.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the CNR project FOE-2019, grant number DBA.AD003.139 and by a grant from the Italian Ministry of Research (MIUR, PON01_01966 'EnerbioChem').

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors would like to thank Gerlind Sulzenbacher from AFMB for the critical discussion on the specificity of the LacS substrate during the 6 months spent by NC at AFMB. The stay of NC at AFMB was financially supported by UniNA and Compagnia di San Paolo, in the frame of Programme STAR. We are grateful to Giovanni del Monaco, Chiara Nobile, and Marco Petruzziello (IBBR-CNR) for administrative and technical assistance.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Broeker, J.; Mechelke, M.; Baudrexl, M.; Mennerich, D.; Hornburg, D.; Mann, M.; Schwarz, H.W.; Liebl, W.; Zverlov, V. The hemicellulose-degrading enzyme system of the thermophilic bacterium Clostridium stercorarium: Comparative characterisation and addition of new hemicellulolytic glycoside hydrolases. *Biotechnol. Biofuels* **2018**, *11*, 229.
2. Lombard, V.; Golaconda, R.H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490–D495.
3. Chen, C.; Dai, L.; Ma, L.; Guo, R. Enzymatic degradation of plant biomass and synthetic polymers. *Nat. Rev. Chem.* **2020**, *4*, 114–126.
4. Strazzulli, A.; Cobucci-Ponzano, B.; Iacono, R.; Giglio, R.; Maurelli, L.; Curci, N.; Schiano-di-Cola, C.; Santangelo, A.; Contursi, P.; Lombard, V.; et al. Discovery of hyperstable carbohydrate-active enzymes through metagenomics of extreme environments. *FEBS J.* **2019**, *287*, 1116–1137.
5. Suleiman, A.; Krüger, A.; Antranikian, G. Biomass-degrading glycoside hydrolases of archaeal origin. *Biotechnol. Biofuels* **2020**, *13*,153.
6. Liu, G.; Qin, Y.; Li, Z.; and Qu, Y. Development of highly efficient, low-cost lignocellulolytic enzyme systems in the post-genomic era. *Biotechnol. Adv.* **2013**, *31*, 962–975.
7. Avanthi, A.; Kumar, S.; Sherpa, K.C.; Banerjee, R. Bioconversion of hemicelluloses of lignocellulosic biomass to ethanol: An attempt to utilize pentose sugars. *Biofuels* **2017**, *8*, 431–444.
8. Zavyalova, A.V.; Rykova, S.V.; Luninab, N.A.; Sushkovaa, V.I.; Yarotskya, S.V.; Berezina, O.V. Plant Polysaccharide Xyloglucan and Enzymes That Hydrolyze It. *Russ. J. Bioorg. Chem.* **2019**, *45*, 845–859.

9. Fry, C.; York, S.; Albersheim, P.; Darvill, A.; Hayashi, T.; Joseleau, J-P.; Kato, Y.; Lorences, P.; Maclachlan, A.; McNeil, M.; Mort, J.; et al. An unambiguous nomenclature for xyloglucan-derived oligosaccharides. *Physiol. Plant.* **1993**, *89*,1–3.
10. Tuomivaara, S.T.; Yao, K.; O'Neill, M.A.; York, W.S. Generation and structural validation of library of diverse xyloglucan-derived oligosaccharides, including an update on xyloglucan nomenclature. *Carbohydr. Res.* **2015**, *402*, 56–66.
11. van der Brink, J.; de Vries, R.P. Fungal enzyme sets for plant polysaccharide degradation. *Appl. Microbiol. Biotechnol.* **2011**, *91*, 1477–1492.
12. Larsbrink, J.; Rogers, T.E.; Hemsworth, G.R.; McKee, L.S.; Tauzin, A.S.; Spadiut, O.; Klinter, S.; Pudlo, N.A.; Urs, K.; Koropatkin, N.M.; et al. A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **2014**, *506*, 498–502.
13. Gardner, J.C. Polysaccharide degradation systems of the saprophytic bacterium *Cellvibrio japonicus*. *World J. Microbiol. Biotechnol.* **2016**, *32*, 121.
14. Ravachol, J.; de Philip, P.; Borne, R.; Mansuelle, P.; Maté, M.J.; Perret, S.; Fierobe, H-P. Mechanisms involved in xyloglucan catabolism by the cellulosome-producing bacterium *Ruminiclostridium cellulolyticum*. *Sci. Rep.* **2016**, *6*, 22770.
15. Attia, M.A.; Brumer, H. Recent structural insights into the enzymology of the ubiquitous plant cell wall glycan xyloglucan. *Curr. Opin. Struct. Biol.* **2016**, 40, 43–53.
16. Ulas, T.; Riemer, A.; Zaparty, M.; Siebers, B.; Schomburg, D. Genome-Scale Reconstruction and Analysis of the Metabolic Network in the Hyperthermophilic Archaeon *Sulfolobus Solfataricus*. *PLoS ONE* **2012**, *7*, e43401.
17. Sakai, H.D.; Norio, K.N. *Saccharolobus caldissimus* gen. nov., sp. nov., a facultatively anaerobic iron-reducing hyperthermophilic archaeon isolated from an acidic terrestrial hot spring, and reclassification of *Sulfolobus solfataricus* as *Saccharolobus solfataricus* comb. nov. and *Sulfolobus shibatae* as *Saccharolobus shibatae* comb. nov. *Int. J. Syst. Evol. Microbiol.* **2018**, *68*, 1271–1278.
18. Wolf, J.; Stark, H.; Fafenrot, K.; Albersmeier, A.; Pham, T.K.; Müller, K.B.; Meyer, B.H.; Hoffmann, L.; Shen, L.; Albaum, S.P.; et al. A systems biology approach reveals major metabolic changes in the thermoacidophilic archaeon *Sulfolobus solfataricus* in response to the carbon source L-fucose versus D-glucose. *Mol. Microbiol.* **2016**, *102*, 882–908.
19. Elferink, M.G.L.; Albers, S.V.; Konings, W.N.; Driessen, A.J.M. Sugar transport in *Sulfolobus solfataricus* is mediated by two families of binding protein-dependent ABC transporters. *Mol. Microbiol.* **2008**, *39*, 1494–1503.
20. Lalithambika, S.; Peterson, L.; Dana, K.; Blum, P. Carbohydrate hydrolysis and transport in the extreme thermoacidophile *Sulfolobus solfataricus*. *Appl. Environ. Microbiol.* **2012**, *78*, 7931–7938.
21. Moracci, M.; Cobucci-Ponzano, B.; Trincone, A.; Fusco, S.; De Rosai, M.; van der Oost, J.; Sensen, C.W.; Charlebois, R.L.; Rossi, M. Identification and Molecular Characterization of the first α-Xylosidase from an Archaeon. *J. Biol. Chem.* **2000**, *275*, 22082–22089.
22. Limauro, D.; Cannio, R.; Fiorentino, G.; Rossi, M.; Bartolucci, S. Identification and molecular characterization of an endoglucanase gene, celS, from the extremely thermophilic archaeon *Sulfolobus solfataricus*. *Extremophiles* **2001**, *5*, 213–219.
23. Moracci, M.; Ciaramella, M.; Rossi, M. Beta-glycosidase from *Sulfolobus solfataricus*. *Methods Enzymol.* **2001**, *330*, 201–215.
24. She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-Weiher, C. C-Y.; Clausen, I.G.; Curtis, B.A.; De Moors, A.; et al. The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 7835–7840.
25. Cobucci-Ponzano, B.; Trincone, A.; Giordano, A.; Rossi, M.; Moracci, M. Identification of an archaeal α-L-fucosidase encoded by an interrupted gene. *J. Biol. Chem.* **2003**, *278*, 14622–14631.
26. Cannio, R.; Di Prizito, N.; Rossi, M.; Morana, A. A xylan-degrading strain of *Sulfolobus solfataricus*: Isolation and characterization of the xylanase activity. *Extremophiles* **2004**, *8*, 117–124.
27. Huang, Y.; Krauss, G.; Cottaz, S.; Driguez, H.; Lipps, G. A highly acid-stable and thermostable endo-beta-glucanase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. *Biochem. J.* **2005**, *385*, 581–588.
28. Brouns, S.J.J.; Smits, N.; Wu, H.; Snijders, A.P.L.; Wright, P.C.; de Vos, W.M.; van der Oost, J. Identification of a Novel α-Galactosidase from the Hyperthermophilic Archaeon *Sulfolobus solfataricus*. *J. Bacteriol.* **2006**, *188*, 2392–2399.
29. Kambourova. M.; Mandeva, R.; Fiume, I.; Maurelli, L.; Rossi, M.; Morana, A. Hydrolysis of xylan at high temperature by co-action of the xylanase from Anoxybacillus flavithermus BC and the beta-xylosidase/alpha-arabinosidase from *Sulfolobus solfataricus* Oalpha. *J. Appl. Microbiol.* **2007,** *102*, 1586–1593.
30. Cobucci-Ponzano, B.; Aurilia, V.; Riccio, G.; Henrissat, B.; Coutinho, P.M.; Strazzulli, A.; Padula, A.; Corsaro, M.M.; Pieretti, G.; Pocsfalvi, G.; et al. A New Archaeal β-Glycosidase from *Sulfolobus solfataricus*. Seeding a novel retaining β-glycan-specific glycoside hydrolase family along with the human non-lysosomal glucosylceramidase GBA2. *J. Biol. Chem.* **2010**, *285*, 20691–20703.
31. Cobucci-Ponzano, B.; Conte, F.; Strazzulli, A.; Capasso, C.; Fiume, I.; Pocsfalvi, G.; Rossi, M.; Moracci, M. The molecular characterization of a novel GH38 α-mannosidase from the crenarchaeon *Sulfolobus solfataricus* revealed its ability in de-mannosylating glycoproteins. *Biochimie* **2010**, *92*, 1895–1907.
32. Honarbakhsh, M.; Villafane, A.A.; Ruhl, I.; Sannino, D.; Bini, E. Development of a thermostable β-glucuronidase-based reporter system for monitoring gene expression in hyperthermophiles. *Biotechnol. Bioeng.* **2012**, *109*, 1181–1186.
33. Briggs, D.C.; Yoshida-Moriguchi, T.; Zheng, T.; Venzke, D.; Anderson, M.E.; Strazzulli, A.; Moracci, M.; Yu, L.; Hohenester, E.; Campbell, K.P. Structural basis of laminin binding to the LARGE glycans on dystroglycan. *Nat. Chem. Biol.* **2016**, *12*, 810–814.
34. Larsbrink, J.; Izumi, A.; Ibatullin, F.M.; Nakhai, A.; Gilbert, H.J.; Davies, G.J.; Brumer, H. Structural and enzymatic characterization of a glycoside hydrolase family 31 α-xylosidase from Cellvibrio japonicus involved in xyloglucan saccharification. *Biochem. J.* **2011**, *436*, 567–580.

35. Kumar, R.; Henrissat, B.; Coutinho, P.M. Intrinsic dynamic behavior of enzyme: Substrate complexes govern the catalytic action of β-galactosidases across clan GH-A. *Sci. Rep.* **2019**, *9*, 10346.
36. Chen, J.M.; Xia, Y.; Wan, H.; Wang, H.; Liu, X. A complete specific cleavage of glucosyl ester linkages of stevioside for preparing steviol with β-galactosidase from *Sulfolobus solfataricus*. *J. Mol. Catal. B Enzym.* **2014**, *105*, 126–131.
37. Bauer, S.; Vasu, P.; Persson, S.; Mort, A.J.; Somerville, C.R. Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11417–11422.
38. Rogowski, A.; Briggs, J.A.; Mortimer, J.C.; Tryfona, T.; Terrapon, N.; Lowe, E.C.; Baslé, A.; Morland, C.; Day, A.M.; Zheng, H.; et al. Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* **2015**, *6*, 7481.
39. Sakurama, H.; Tsutsumi, E.; Ashida, H.; Katayama, T.; Yamamoto, K.; Kumagai, H. Differences in the Substrate Specificities and Active-Site Structures of Two α-L-Fucosidases (Glycoside Hydrolase Family 29) from *Bacteroides thetaiotaomicron*. *Biosci. Biotechnol. Biochem.* **2012**, *76*, 1022–1024.
40. Paper, J.M.; Scott-Craig, J.S.; Cavalier, D.; Faik, A.; Wiemels, R.E.; Borrusch, M.S.; Bongers, M.; Walton, J.D. α-Fucosidases with different substrate specificities from two species of *Fusarium*. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 5371–5380.
41. Vecchio, G.; Parascandolo, A.; Allocca, C.; Ugolini, C.; Basolo, F.; Moracci, M.; Strazzulli, A.; Cobucci-Ponzano, B.; Laukkanen, M.; Castellone, M.; et al. Human α-L-fucosidase-1 attenuates the invasive properties of thyroid cancer. *Oncotarget* **2017**, *8*, 27075–27092.
42. Ashida, H.; Miyake, A.; Kiyohara, M.; Wada, J.; Yoshida, E.; Kumagai, H.; Katayama, T.; Yamamoto, K. Two distinct α-L-fucosidases from *Bifidobacterium bifidum* are essential for the utilization of fucosylated milk oligosaccharides and glycoconjugates. *Glycobiology* **2009**, *19*, 1010–1017.
43. Cao, H.; Walton, J.D.; Brumm, P.; Phillips, G.N. Jr. Crystal structure of α-xylosidase from *Aspergillus niger* in complex with a hydrolyzed xyloglucan product and new insights in accurately predicting substrate specificities of GH31 family glycosidases. *ACS Sustain. Chem. Eng.* **2020**, *8*, 2540–2547.
44. Chaillou, S.; Lokman, B.C.; Leer, R.J.; Posthuma, C.; Postma, P.W.; Pouwels, P.H. Cloning, sequence analysis, and characterization of the genes involved in isoprimeverose metabolism in *Lactobacillus pentosus*. *J. Bacteriol.* **1998**, *180*, 2312–2320.
45. Matsuzawa, T.; Kameyama, A.; Yaoi, K. Identification and characterization of α-xylosidase involved in xyloglucan degradation in *Aspergillus oryzae*. *Appl. Microbiol. Biotechnol.* **2019**, *104*, 201–210.
46. Okuyama, M.; Kang, M-S.; Mitsuishi, Y.; Mori, H.; Kimura, A. Substrate recognition of *Escherichia coli* YicI (α-Xylosidase). *J. Appl. Glycosci.* **2008**, *55*, 111–118.
47. Sampedro, J.; Sieiro, C.; Revilla, G.; Gonzalez-Villa, T.; Zarra, I. Cloning and expression pattern of a gene encoding an alpha-xylosidase active against xyloglucan oligosaccharides from *Arabidopsis*. *Plant Physiol.* **2001**, *126*, 910–920.
48. Bradford, M.M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **1976**, 72, 248–254.
49. Wang, Q.C.; Zhao, X.; Pu, J.H.; Luan, X.H. Influences of acid reaction and hydrolytic conditions on monosaccharide composition analysis of acid, neutral and basic polysaccharides. *Carbohydr. Polym.* **2016**, *143*, 296–300.

4.3. Conclusion

Enzymes with higher operational stability are needed in order to improve the bioconversion of lignocellulose biomasses. Moreover, considering the importance of hemicellulose in the plant cell wall structure, hemicellulases in combination with cellulolytic enzymes play a crucial role in deconstructing lignocellulose feedstock. The three enzymes from *S. solfataricus* described in this work were already biochemically characterized in detail. However, a poor characterization was previously performed on natural substrates, leaving unexplored their potential as biocatalysts in the hydrolysis of hemicelluloses. In this study, their ability to hydrolyze XGOs was investigated, showing excellent operational stability. In particular, the GH29 remove all α-fucose residues from the substrate tested. At the same time, the GH1 and GH31 displayed significant synergism between them. The data obtained demonstrated the importance of testing the enzyme activity on natural substrates or on substrates with industrial interests in order to explore their potential as biocatalysts for biotechnological application.

CHAPTER 5

5. GH109 and blood Group A conversion

5.1. Overview

Due to the many functions performed by GHs, they are largely employed for industrial and biotechnological purposes and can be also used for the development of new and meaningful applications. As mentioned above, carbohydrates are involved in many essential functions in all living beings. In humans, carbohydrates are also involved in the self- or non-self-recognition by the immunity system. One of such examples is the red blood cell (RBC) recognition system, which, depends on the composition of oligosaccharides on the erythrocytes' surface. Based on the antigen present on the RBCs surface, the blood is classified into different groups, according to the AB0 classification [160]. In the field of blood transfusions, incorrect administration of blood group to an individual could lead to fatal consequences. In these regards, it was demonstrated that some GHs could be helpful in the conversion of A and B type blood to produce the blood universal donor group 0 [161]. Some GHs belonging to the GH109 family are of particular interest for this biotechnological application due to their ability to convert blood group A in group 0 [107, 162]. Improving this bioconversion would bring great benefits in the clinical field and in the global blood industries. The sales of global blood products were estimate of 31.8 billion USD in 2015 and 33.5 billion USD in 2016. The market is expected to grow at a five-year CAGR of 4.9% from 2016 to 2021 [163]. The discovery of enzymes capable of operating in this bioconversion of blood and its development on a large scale induced the development of companies involved in this biotechnology, such as Velico Medical in the USA, aimed for this purpose. Moreover, find novel enzymes that improve the methodology could lead to the massive production and sale of these enzymes. This chapter describes the enzymatic characterization of three novel GH109 enzymes and the detailed study of their potential for blood bioconversion. The study has been done in collaboration with the company Novozymes, which identified the three enzymes and gave us enough amounts of recombinant enzymes to perform the experiments.

5.2. Introduction

According to the Human Health Organization, the worldwide blood supplies demand is constantly increasing, and over the years, several attempts have been placed to stabilize these stocks. Enzymatic conversion technology of red blood cells (RBC) A-type and B-type to universal donor 0-type (or H-type) is considered a valuable alternative to overcome the low-blood supplies issue [161]. 0-type RBC is the most

common in world population distribution and can be transferred to group A, B and AB subjects. Thus, in case of limited supplies or emergencies, in which the subject's RBCs group is not clear or unknown, a sufficient amount of 0-type blood supplies in blood banks became indispensable [161]. The total global blood supply was estimated to be around 272 million units. However, the total global demand in 2017 was approximately 303 million units, a shortfall of around 30 million units [164].

The AB0 classification system for RBCs was described for the first time by Karl Landsteiner in 1900 [160]. This classification is based on the presence or absence of specific antigens on the RBCs surfaces, represented by oligosaccharides linked to the erythrocyte cell membranes' glycolipids and glycoproteins. The antigenic determinants' base structure is a fucosyl galactose (Fuc-α-1,2-β-Gal) that characterized the 0-type group. The addition of extra terminal monosaccharides forms the A-type and B-type antigen, in which the immunodominant sugar moiety is an α-1,3-linked galactose (Gal) for group B and α-1,3-linked *N*-acetylgalactosamine (GalNAc) for group A (Fig. 12) [161].

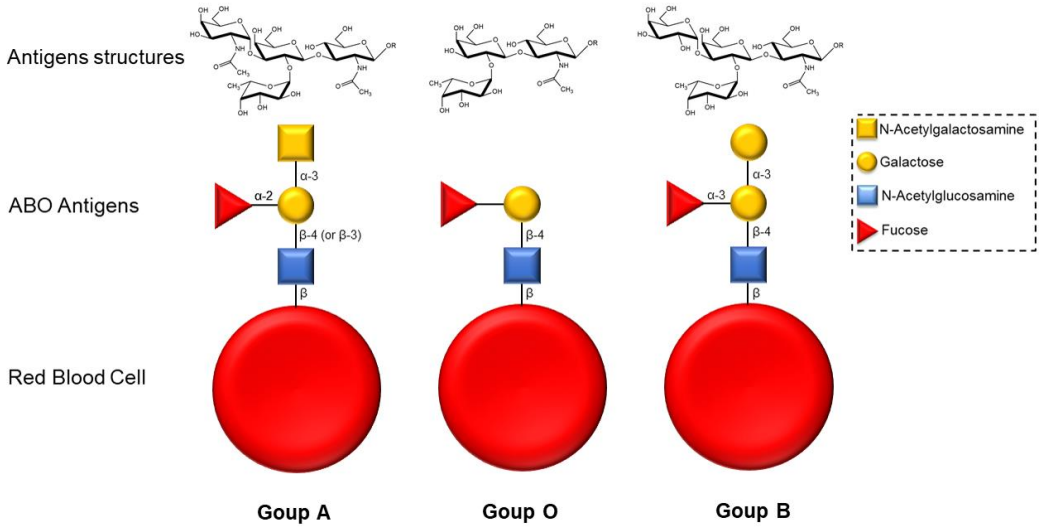


Figure 12: ABO antigens representation

Group A could be distinguished in two subgroups, namely A_1, A_2, based on the structure of the oligosaccharides. The structure of type A (Fuc-α-1,2-βGal-1,3-α-GalNAc) could be further decorated with a GlcNAc moiety in β-1,4 and in β-1,3 glycosidic bond, for A_1 and A_2 respectively (Fig 12). The different structure of the two A subgroups does not

influence the immunogenic recognition of group A, however the subjects having a structure of type A_1 present a number of A-antigen sites 6 fold higher than those having the A_2 structure, 1.5 versus 0.25 million respectively [162].
People with A-type RBC group have anti-B antibodies, while individuals with B-type have anti-A antibodies. Transfusion of A-type blood in B-type recipients, and *vice versa*, would lead to agglutination of RBCs and subsequent haemolysis. Individuals with blood group AB have both antigens A-type and B-type determinants and by consequence, they do not have antibody against blood antigens. People with 0-type blood have both anti-A and anti-B antibodies, and they can receive only 0-type RBCs. 0-type blood lack of both immunodominant monosaccharides of A and B groups, and neither anti-A nor anti-B antibodies lead to agglutination. Considering this immunogenicity, due to a single monosaccharide on the RBCs surface, the selective removal of the antigenic determinants via enzymatic hydrolysis can produce a 0-type universal donor. The conversion of A-type or B-type antigen in 0-type via enzymatic cleavage, does not affect the Rh status. Accordingly, the enzymatic conversion of RBC group A or B Rh + would produce RBC group 0 Rh +, whereas RBCs A or B Rh - would produce only 0-type RBC Rh -.
The first attempt to convert RBCs antigenicity via enzymatic hydrolysis was performed in 1980 by Goldstein and co-workers using α-galactosidases from coffee bean to remove galactose on RBC B-type surfaces. They obtained the conversion, and experiments on gibbon and even human volunteers confirmed that the transformed RBCs exhibited an average half-life and were well-tolerated. However, the conversion required large amounts of enzymes and low pH, making the procedure ineffective for the integrity of erythrocyte and also in terms of costs [165-166]. In this frame, Liu and colleagues in 2007 find a GH109 with α-*N*-acetylgalactosaminidase activity and a GH110 with α-galactosidase activity, capable of removing the immunodominant monosaccharides on A-type and B-type, respectively. Although GH110 convert the group B RBCs efficiently, the conversion of the A group by the GH109 required significant amounts of enzyme and particular buffer conditions [107]. Other GH109 enzymes of this family were discovered and tested on A-type RBCs, even improving the efficiency of bio-conversion, confirming that GH109 enzymes can be suitable for this biotechnological application [161,168]. To date, the best method with low enzymatic loading to convert group A is based on the utilization of two different enzymatic activities. Rahfeld et al. developed an assay to cleave the GalNAc A-type antigen in two steps: by using a GalNAc

acetylase that removes the N-acetyl group from GalNAc, and then an α-galactosaminidase in order to remove the galactosamine left by the previous enzyme (Fig. 13) [169].

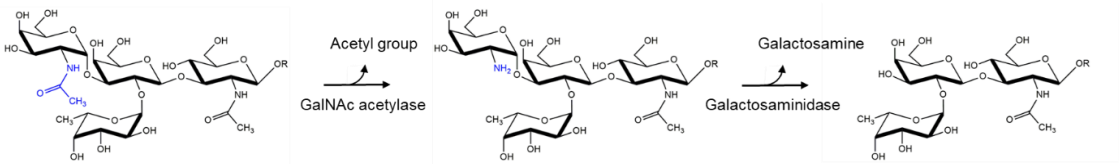


Figure 13: Double step A-type antigen removal by GalNAc acetylase and Galactosaminidase

Finding new enzymes that improve this biotechnology is needed to lower the costs and develop the universal blood production, meeting the demand for worldwide blood supplies. In this work, in collaboration with Novozymes company, we characterized three novels' GHs of family GH109 and tested their bio-conversion power on group A RBCs.

5.3. Materials and Methods

GH109 enzymes (NAg68; NAg69; NAg71) were provided by Novozymes (Bagsværd, Copenhagen, Denmark). Blood samples and Anti-A serum (ImmuClone Anti-A, Birma-1, BIOSCOT®) were provided by the department of molecular medicine and medical biotechnology, University of Naples "Federico II" (Naples, Italy). 4NP-substrates and Blood Group-A trisaccharide was purchased from Biosynth-Carbosynth® (St. Gallen, Switzerland).

5.3.1. In silico analysis

Multisequence alignment of the GH109 conserved motif was performed by Clustal Omega tools (Clustal Omega < Multiple Sequence Alignment < EMBL-EBI).

5.3.2. Molecular mass determination

Native molecular mass of NAg68, NAg69, NAg71 were examined by gel filtration using AKTA prime FPLC equipped with Superdex 10/300 GL (Cytiva). The run was performed in buffer Hepes 50 mM NaCl 100 mM pH 8.0, at 0.7 mL/min. The calibration line was build using protein marker carbonic anhydrase (29000 Da), bovine serum albumin (66103 Da) alcohol dehydrogenase (150000 Da), β-amylase (200000 Da), apoferritin (443000 Da) and thyroglobulin (669000 Da).

5.3.3. Activity assay

In a final volume of 400 µL, the standard assay of GH109 enzymes (7.5 µg of NAg69, 5 µg of NAg68, and 2.5 µg of NAg71) was performed in 50 mM sodium phosphate pH 8.0 on 3 mM of 4NP-βGalNAc at 40 °C

for NAg69 and 45° for NAg71 and NAg68. The 4NP release (the mM coefficient extinction in the assay condition was 13.2 mM^{-1} cm^{-1}) was monitored continuously at 420 nm by using double bean spectrophotometer Cary 100 UV. One unit of enzymatic activity was defined as the amount of enzyme catalyzing the hydrolysis of 1 µmol os substrate in 1 min at the condition described.

5.3.4. Temperature and pH dependence

Optimal pH and buffer saline conditions were determined by using sodium phosphate buffer in the pH range 6.0 - 8.0 and sodium borate buffer in the range of 8.0 - 9.0. Effect of temperature on enzymes activity was evaluated in the range of 20-50 °C in sodium phosphate buffer pH 8.0 on 3 mM of 4NP-βGalNAc as substrate.

5.3.5. Effect of additives

Assays were carried out in the enzymes standard condition by adding 2 mM of each additives (NAD^+, EDTA, $CuCl_2$, $CaCl_2$, $MgSO_4$, $ZnSO_4$, $FaCl_2$) in the reaction mixture. Control referred to the assay in standard condition without additives.

5.3.6. Kinetic parameter determination

Steady-state kinetic constants of the enzymes were measured in the standard conditions on various concentration of 4NP-α- and 4NP-βGalNAc. NAg69: 0.01 - 2 mM of 4NP-αGalNAc and 0.05 - 5 mM 4NP-βGalNAc. NAg68: 0.01 - 2 mM of 4NP-αGalNAc and 0.05 - 5 mM of 4NP-βGalNAc. NAg71: 0.01 - 2 mM of 4NP-αGalNAc and 0.25 - 2mM of 4NP-βGalNAc. The data were analysed with GraphPad Prism 8 program, by nonlinear regression using Michaelis-Menten equation and linear regression using Lineweaver-Burk equation for kinetic parameters of NAg71 on 4NP-βGalNAc.

5.3.7. Thermal stability

The thermal stability of NAg68, NAg69 and NAg71 was evaluated at 37°C by testing their activities in their standards conditions after incubation of the enzyme solution at 37°C for different times.

5.3.8. Activity on Blood Group A trisaccharide

The activity of the NAg68, NAg69 and NAg71 on A-type RBC antigen trisaccharide as substrate was evaluated at 37 °C, by using 300 µg of substrate and 50 µg of enzymes in sodium phosphate buffer pH 8.0. An assay mixture without enzymes was used as control. All reactions were stopped by freezing in dry ice after incubation of 10 minutes. The specific activity was evaluated by using 2 µg of each enzyme on 2 mM

of substrate, in sodium phosphate buffer pH 8.0 at 37 °C. One unit of enzymatic activity was defined as the amount of enzyme catalyzing the hydrolysis of 1 µmol of substrate in 1 min at the condition described. The reactions were stopped in dry ice after 1 min. The reaction products were evaluated and quantified by HAEPC-PAD analysis.

5.3.9. HPAEC-PAD analysis

The analysis was performed by using HPAEC-PAD Dionex system (ICS 3000, Waltham, MA, USA) equipped with PA-100 column (Dionex, Sunnyvale, CA, USA) in a linear gradient of 11 mM NaOH. GalNAc was used as standard to calibrate the retention time and to build a calibration line at concentration of 0.5 nmol 25 μL^{-1}, 1.0 nmol 25 μL^{-1}, and 2.0 nmol 25 μL^{-1} with fucose as internal standard at concentration of 1 nmol 25 μL^{-1}. 25 µL of diluted stopped reactions supplemented with 1 nmol of fucose were injected for HAEPC-PAD analysis. Moreover, to confirm the substrate and GalNAc signals in the HPAEC-PAD profile, a little amount of substrate or GalNAc were added to different aliquot of the reactions and analysed separately.

5.3.10. Activity on RBCs

The whole blood group A (provided by the department of molecular medicine and medical biotechnology, University of Naples 'Federico II') was centrifuged at 1000 g for 5 min at room temperature and RBCs were separated from plasma. RBCs were washed three times in PBS buffer (150 mM NaCl, 50 mM sodium phosphate) pH 7.4 and resuspended at 10 % haematocrit in the same buffer. In this way the cells are not subjected to osmotic stress or pH changes, ensuring their integrity during enzymatic tests. For the assay in the presence of dextran, RBCs were resuspended at 10 % haematocrit in PBS pH 7.4, supplemented with 300 mg mL^{-1} of dextran 40 KDa. In a final volume of 200 µL various enzymes amounts (0.1, 0.2, 1, 2, 4, 8 and 16 µg) were added to the resuspended RBCs. The reaction mixes were incubated at 37 °C for 1 h under shaking. The reaction was stopped by cooling in ice and centrifuge at 1000 g for 5 min to separate RBCs from surnatant (containing the enzymes).

5.3.11. Agglutination assays

After the reaction, RBC were washed three times with PBS pH 7.4 and then resuspended in 180 µL of the same buffer. 10 µL of resuspended RBCs were mixed with 2 µL of Anti-A serum.

5.3.12. FACS analysis

Enzyme-treated RBCs were washed twice with PBS and BSA pH 7.4 and 1% haematocrit enzymatically converted RBCs were treated with 1/100 APC–anti-A antibody (Alexa Fluor 647 mouse anti-human blood group A: cat. no. 565384 (BD Pharmingen)) for 30min at room temperature. Flow cytometry was performed after reconstitution into 1× PBS pH7.4 (1% haematocrit) on a FACSCantoII flow cytometer using FACSDiva v 9.0. The data were refined by FlowJ v 10. RBCs were identified based on the FSC and SSC parameters and a gate was applied to select single cells.

5.4. Results and discussion

5.4.1. Enzymes' characterization

The enzymes analysed in this work, NAg68, NAg69 and NAg71 belong to the genus Alkalimonas, Rheinhera and Amycolatopsis, respectively. Multi sequences alignment with the other GH109 characterized revealed that these enzymes share the consensus motif GGHGG, containing the catalytic acid-base histidine (Fig. 14)

Figure 14: Clustal Omega multisequence alignment

NAg68, NAg69 and NAg71 had a molecular mass of 71.7 KDa, 74.2 KDa and 56.02 KDa, respectively. The enzymes dependence on temperature and pH, and the effect of additives, were evaluated using 4NP-βGalNAc substrate, as reported in Materials and Methods. NAG68 was optimally active in sodium phosphate buffer at pH 8.0 and 45 °C on 4NP-βGalNAc (Fig 15) and exhibited a specific activity in standard condition of 11.6 U mg^{-1} and 7 U mg^{-1} for 4NP-α- and βGalNAc, respectively. NAg69 showed an optimal activity temperature of 40 °C and, similarly to the previous enzyme, an optimum pH of 8.0 (Fig 15). The measured specific activities were of 12 U mg^{-1} on 4NP-αGalNAc

and 2.7 U mg^{-1} on the β-anomer. NAg71, like the other two enzymes, had a pH optimum of 8.0 and is optimally active at 45 °C, like the NAg68 (Fig 15). The enzyme had a specific activity of 11 U mg^{-1} on 4NP-α-GalNAc and 7.2 U mg^{-1} on the β-anomer.

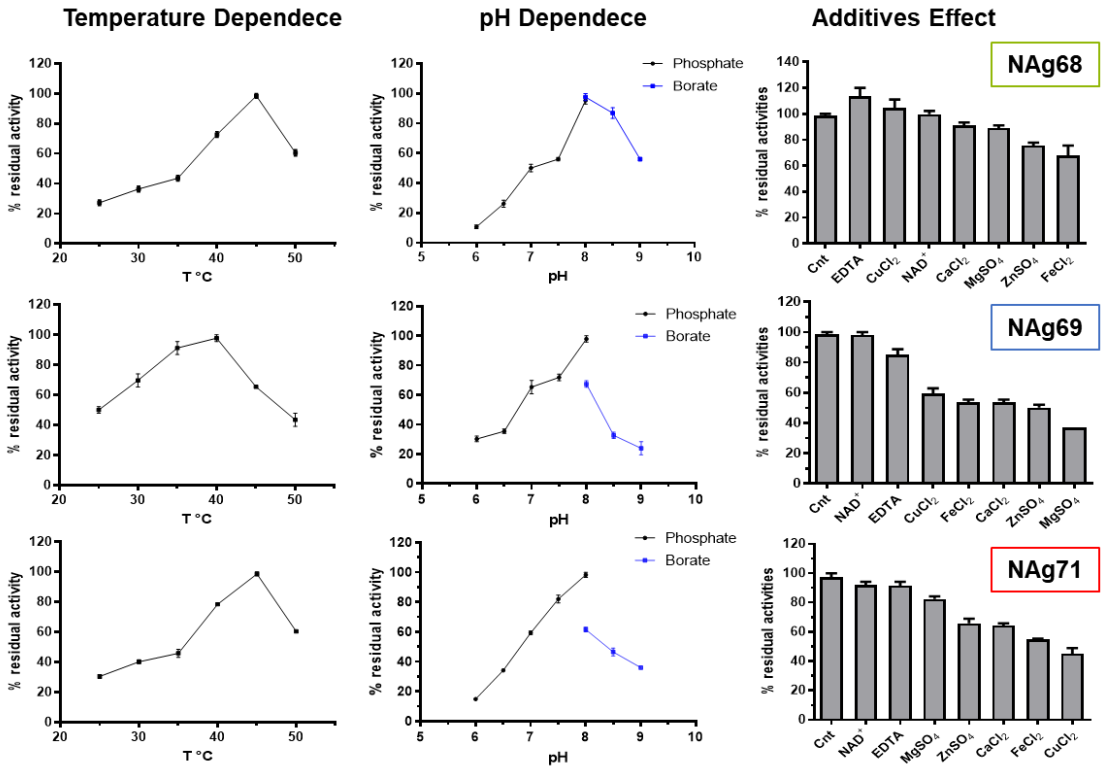


Figure 15: Characterization of GH109 enzymes: Temperature and pH dependences, and additives effect of NAg68; NAg69 and NAg71.

The tree enzymes did not require exogenous addition of NAD^+ . Their activity in the presence of 2 mM of the cofactor was similar to the standard assays in its absence (Fig. 15). These data are in accordance with the other mesophilic GH109 enzymes in which the cofactor is tightly bound to the active site of the enzymes, and keeped once purified in recombinant form. The presence of NAD^+ in the structure has been experimentally demonstrated for the first characterized GH109 enzyme from *Elizabethkingia meningoseptica* (NagA) and by the solved structure of two GH109 enzymes from *Akkermansia muciniphila* [107,108]. EDTA had no negative effect on the enzymes and exogenous metal ions did not enhance the enzymes' activities, indicating that they did not require metal ions for their catalysis. EDTA had a quite activating effect on NAg68, while all metal ions had negative effects on NAg69 and NAG71 (Fig. 15). The three GH109 were tested

on different aryl substrates (4NP-α/β-glucopyranoside, -galactopyranoside, and -GalNAc) showing higher and measurable activity only on aryl- α/β-GalNAc, on which the kinetic constants have been determined.

Table 1: Kinetic parameters of NAg68, NAg69 and NAg71

Enzymes	Substrates	K_M (mM)	k_{cat} (s^{-1})	k_{cat} K_M^{-1} (s^{-1} mM^{-1})
NAg68	4NP-αGalNAc	0.05 ± 0.004	13.9 ± 0.2	268.4 ± 0.1
	4NP-βGalNAc	0.23 ± 0.05	8.5 ± 0.3	36.7 ± 0.3
NAg69	4NP-αGalNAc	0.01 ± 0.004	15 ± 0.6	1085 ± 0.3
	4NP-βGalNAc	0.11 ± 0.02	3.3 ± 0.03	30.6 ± 0.2
NAg71	4NP-αGalNAc	0.03 ± 0.004	20.6 ± 0.2	603.7 ± 0.2
	4NP-βGalNAc	0.31 ± 0.06	14.1 ± 0.03	45.4 ± 0.2

The kinetic parameters reported in Table 2 show that 4NP-αGalNAc was the preferred substrate for all the three enzymes. NAg69 had different k_{cat} for the two substrates but of the same order of magnitude (8.5 and 13.9 s^{-1} for 4NP-β- and -αGalNAc, respectively). The enzyme showed a K_M of 0.05 mM for 4NP-αGalNAc, almost five-fold lower than that of the other anomer (0.23 mM), resulting in a higher specificity constant for the α-substrate (Table 2). Even NAg69 showed different kinetic parameters for the substrate tested. The K_M for 4NP-αGalNAc is ten-fold lower compared to that on 4NP-βGalNAc (0.01 and 0.11 mM, respectively), the k_{cat} is higher for the α-anomer (15 s^{-1}) (Table 2) and the enzyme exhibited a very high specificity constant of 1085 s^{-1} mM^{-1} for this substrate. NAg71, like NAg69, show a K_M for the 4NP-αGalNAc ten-fold lower than the β-substrate (0.03 and 0.31 mM, respectively) and exhibited a k_{cat} of the same order of magnitude for the two substrates. As a consequence, the specificity constant for the 4NP-αGalNAc is much higher as shown in Table 2. All the three enzymes displayed specific activities on 4NP-αGalNAc comparable to those of NagA (12 U mg^{-1}), the first GH109 identified and active on type-A RBCs [107]. However, compared to the specificity constant of NagA for this substrate (127.6 s^{-1} mM^{-1}), the values of k_{cat}/K_M were 2.1-fold, 8.5-fold, and 4.7-fold higher for NAg68, NAg69, and NAg71, respectively. More in general, compared to the specificity constants available for the other characterized enzymes of family GH109, the value of k_{cat}/K_M of the GH109s in this work are the highest reported so far (Fig. 16).

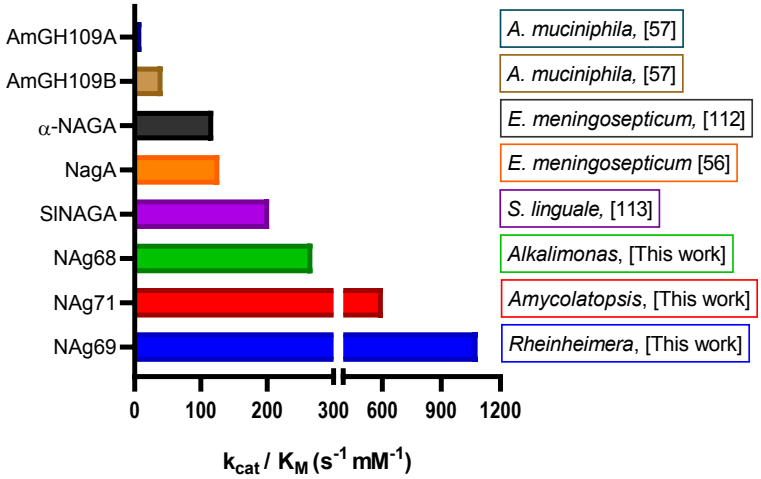


Figure 16: Specificity constants comparison of characterized GH109 on 4NP-αGalNAc

On the basis of their specificity, the three novel GH109 characterized in this work were further characterized to test the possible application on RBCs conversion.

5.4.2. Activity on Blood group A trisaccharide

To evaluate the activity of the three enzymes on a more complex substrate and their potential as biocatalysts in the conversion of A-type RBC we tested their activity on the blood group A trisaccharide (TrBgA). The sugar is the antigenic determinant of type-A RBC and consists in a α-1,3-linked N-acetylgalactosamine to the antigenic base structure of group 0 the Fuc-α-1,2-Gal-β, as shown in figure 17.

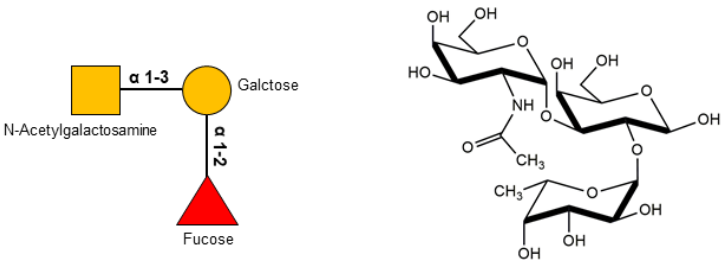


Figure 17: Blood Group A trisaccharide structure representation

The assays were performed at 37 °C in sodium phosphate buffer pH 8.0, as described in Materials and Methods. In this condition the enzymes showed specific activities on 4NP- βGalNAc of 4.5, 2.5 and 5.6 U mg^{-1} and on 4NP-αGalNAc of 7.5, 11.4 and 8.6 U mg^{-1} for NAg68, NAg69, and NAg71, respectively. Before incubating the enzymes with the TrBgA substrate, their thermal stability at 37 °C was evaluated.

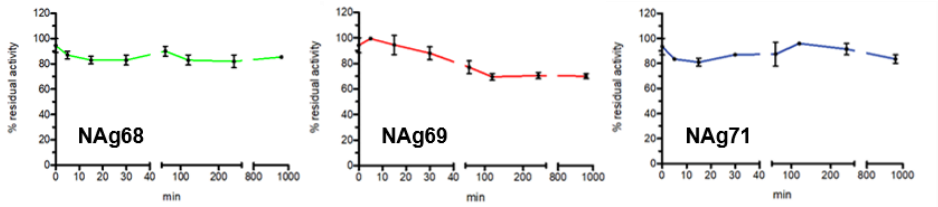


Figure 18: Thermal stability

NAg68 and NAg71 conserved almost 80% of activity, while NAg69 70% of activity after 16 h of incubation at 37 °C, as Figure 13 displays. The reactions of hydrolysis on TrBgA were performed in 10 min on 300 µg of the substrate, and the sugars released were separated and analysed by HPAEC-PAD chromatography .

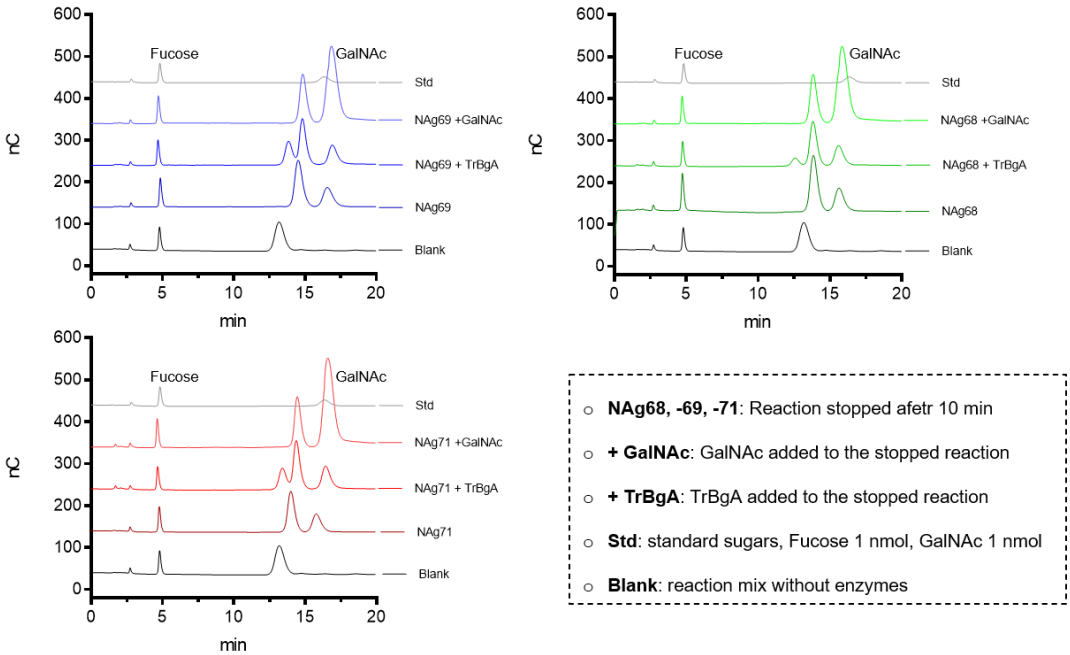


Figure 19: HPAEC PAD chromatograms. Blu lines: NAg68. Green lines: NAg68. Red lines: NAg71

Figure 19 shows the HPAEC-PAD chromatographic profiles of the three enzymes reactions, which are all very similar. The three enzymes can release GalNAc from Type-A RBC trisaccharide. Using the NAg69 chromatograms as a model (Fig. 19), we can see that in the blank profile, the substrate's peak has a retention time of 12.6 min. In the reaction profile (NAg69), on the other hand, the substrate peak is not detectable anymore, but we observed two peaks with retention times of 13.4 and 15.2 min. According to the retention times of the standards (Std), the peak at 15.2 min should correspond to GalNAc, while, by exclusion, the peak at 13.4 should be that of the second reaction

product, the α-*L*-fuc-(1,2)-gal disaccharide. Observing the elution profile of the reaction supplemented with the substrate (NAg69 + TrBgA), we observed the appearance of a peak at 12.6 min, a retention time identical to that of the substrate. More, when GalNAc was added to the stopped reaction (NAg69 + GalNAc), we observe only an increase in the peak area at 15.2 min, corresponding to the peak of GalNAc. These analyses give us a clear indication that the enzymes, after 10 minutes of incubation under the conditions used, convert the Type-A RBC antigen trisaccharide into the related products, GalNAc and α-*L*-fuc-(1,2)-gal disaccharide. The amounts of GalNAc released from 300 µg of TrBgA after 10 min of incubation at 37 °C are 144 ± 4, 124 ± 3, and 135 ± 5 µg by NAg68, NAg69, and NAg71, respectively. Moreover, the specific activities of the three enzymes, calculated on 2 mM substrate, were 20 ± 1, 16 ± 0.3, and 14 ± 0.3 U mg^{-1} for NAg68, NAg69, and NAg71, respectively. Such activities are higher if compared to the enzymes' activity on the aryl substrates, confirming their potentiality for the hydrolysis of GalNAc on RBC antigens.

5.4.3. A-type RBCs conversion

To test if the enzymes were effective on RCB cells, erythrocyte from fresh whole blood were separated from plasma and resuspended in a PBS solution at physiological pH 7.4. Moreover, considering the generally poor association of macromolecules on the cell surfaces due to the repulsion of two hydrophilic components, we tested the activity of the enzymes also in presence of dextran 40 KDa as macromolecular crowder [170]. This polymer forms a bio gel in solution, decreasing the reaction volume available to the enzymes and bringing them closer to erythrocyte surface. Thereby, the enzyme concentration increases around the cell surface possibly improving the cleavage performance, as already shown for other enzymes cleaving RBCs antigens [169,170]. In the preliminary assays on resuspended RBCs at 10 % haematocrit from four different donors, different concentrations of the enzymes were used. The detection to evaluate the conversion was performed through agglutination assays by adding Anti-A serum to the treated RBCs as reported the example in Figure 20.

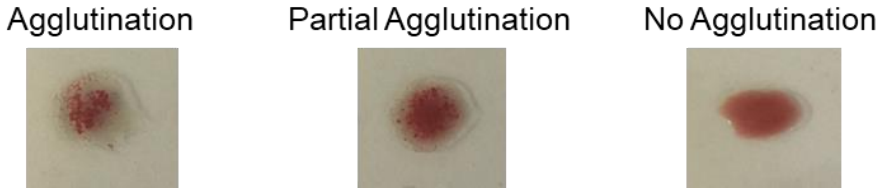


Figure 20: Example of agglutination assays by adding Anti-A serum to RBCs

If the enzymes fully convert the RBCs, agglutination was not observed. Otherwise, we observed agglutination or partial agglutination if the RBCs were not converted or not fully converted.

Table 7: Agglutination assays of RBCs from five different donors treated with different concentration of the three enzymes in presence and in absence of macromolecular crowder.

Enzyme	Sample	0.5 µg/mL	1 µg/mL	2 µg/mL	5 µg/mL	10 µg/mL	20 µg/mL	40 µg/mL	80 µg/mL
NAg68	1	-	-	*	-	-	-	-	-
	2	-	-	*	-	-	-	-	-
	3	-	-	*	-	-	-	-	-
	4	-	-	*	-	-	-	-	-
NAg69	1	-	-	*	-	-	-	-	-
	2	-	-	*	-	-	-	-	-
	3	-	-	*	-	-	-	-	-
	4	-	-	*	-	-	-	-	-
NAg71	1	-	-	*	-	-	-	+	+
	2	-	-	-	-	-	-	+	+
	3	-	-	-	-	-	-	-	+
	4	-	-	-	-	-	-	+	+
NAg68 + dextran	1	-	-	*	+	+	+	+	+
	2	-	-	*	+	+	+	+	+
	3	-	-	*	-	+	+	+	+
	4	-	-/+	*	+	+	+	+	+
NAg69 + dextran	1	-	-	*	-	+	+	+	+
	2	-	-	*	+	+	+	+	+
	3	-	-	*	-	-	+	+	+
	4	-	-/+	*	+	+	+	+	+
NAg71 + dextran	1	+	+	*	+	+	+	+	+
	2	+	+	+	+	+	+	+	+
	3	-	-/+	+	+	+	+	+	+
	4	+	+	+	+	+	+	+	+

- : agglutination + : no agglutination - /+ : Partial agglutination * : No tested condition

Table 7 shows the results of the agglutination tests performed on the treated RBCs (trRBCs) from the four different donors with increasing concentration of the three enzymes in the presence and in absence of macromolecular crowder. The presence of dextran in the assays clearly enhanced the activity of the enzymes on RBCs. Indeed, NAg68 and NAg69 only in the presence of the polymer were able to produce RBCs that not agglutinated by adding Anti-A serum. Remarkably, the concentration of NAg71 needed to convert type A RBCs was markedly reduced in the presence of the macromolecular crowder. As showed in Table 7, the minimum enzymatic concentration (MEC) for full

conversion was not the same in all samples, and it depends on the different donors. For example, in the presence of dextran, to convert RBCs of donors 3, a higher concentration of the three enzymes is always required. Instead, for the other donors, 5 μg mL^{-1} for NAg68 and NAg69 and 0.5 μg mL^{-1} of NAg71 were already enough to fully convert RBCs type A. This difference is most likely due to the subgroup of A-type RBCs, A_1 and A_2. Indeed, RBC with A_1 structure show much more antigens on the surfaces than A_2-type, as explained in the introduction. These data were coherent with those reported by the other GH109 enzymes active on erythrocytes, which required different MEC to convert A_1- and A_2-type RBCs [161]. Based on this observation, we will refer to the MEC needed to convert RBCs to the enzymes' concentration where the trRBCs do not exhibit agglutination in all samples. Accordingly, MECs in the presence of dextran for NAg68, NAg69 and NAg71 were 10, 20 and 2 μg mL^{-1}, respectively.
In order to validate the data from agglutination tests and to estimate the conversion power of the three enzymes, some trRBCs samples by donor 4 were analysed by fluorescence-activated cell sorting (FACS).

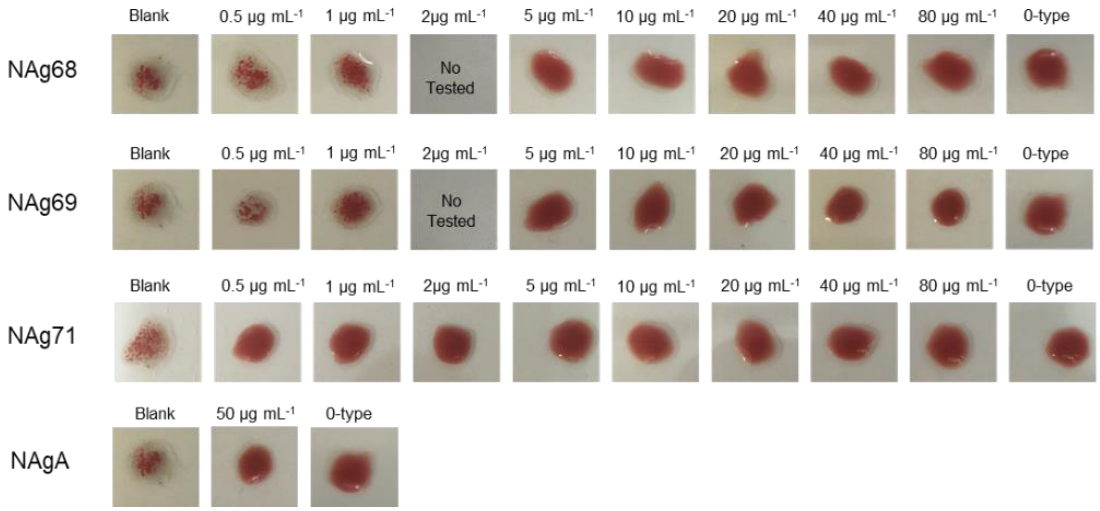


Figure 21: Agglutination tests on treated RBCs from donor 4

Figure 21 shows the agglutination tests of trRBCs from donor 4 by the three enzymes in the presence of dextran. Moreover, we also tested the NagA from *E. meningosepticum* on the same RBCs, as a positive control. The blank tests refer to untreated type-A RBCs, in which strong agglutinations were observed when anti-A serum was added. As a control, the type 0 blood was treated with anti-A serum, and no agglutination was observed. RBCs treated with NAg71 did not show agglutination in all concentration tested. While for both NAg69 and

NAg68 the MEC needed to fully convert this sample were 5 μg mL^{-1}. However, for the latter enzyme it is possible to observe a partial agglutination on the RBCs treated with a concentration as low as 1 μg mL^{-1}. NagA confirmed its capability to convert RBCs group A.

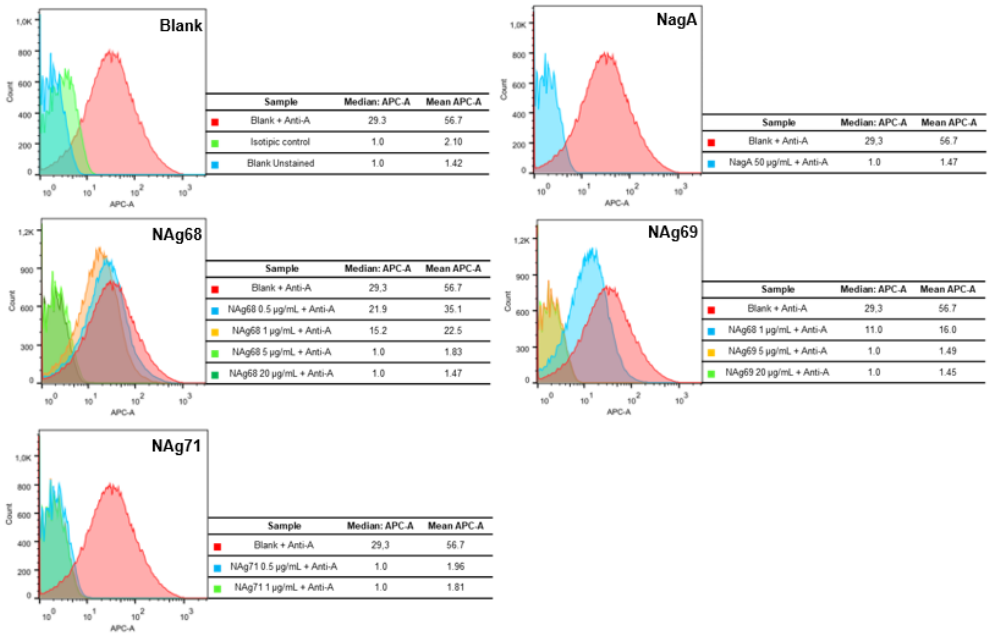


Figure 22: FACS analyses

To perform the FACS analysis we selected the samples that did not showed agglutination at the MEC and one sample at a higher concentration for the three enzymes (1 μg mL^{-1} for NAg71 and 20 μg mL^{-1} for NAg68 and NAg69). The samples that showed a partial agglutination and one sample treated showing agglutination (NAg68 0.5 μg mL^{-1}) were also selected as suitable controls. More, the analysis by FACS was performed even on a blank sample and on the converted RBCs by NagA. Figure 22 shows the FACS analysis of the samples selected treated with APC-labelled anti-A antibody. The blank graph shows the non-treated RBCs stained and unstained with anti-A having a mean APC-A of 56.7 and 1.42, respectively. These two values had been taken as a reference to indicate the maximum point where the Anti-A completely binds the RBCs and the point where there was no antigenic recognition (Blank + Anti-A and Blank unstained respectively in figure 22). Thus, the positive control NagA 50 μg mL^{-1} completely converted group A RBCs with a mean APC-A value of 1.45 (99.9% of conversion), as well as NAg68 at 5 and 20 μg mL^{-1} (99.2 and 99.9%,

respectively), NAg69 at 5 and 20 μg mL^{-1} (99.9 and 99.9%, respectively) and NAg71 at 0.5 and 1 μg mL^{-1} (99% and 99.2%, respectively). The analysis confirmed the partial agglutination of the sample NAg69 and NAg68 at 1 μg mL^{-1}, with a conversion of 61.5 % and 73.5 %. Instead, the sample NAg68 at 0.5 μg mL^{-1} showed a conversion of 39 %, indicating that this concentration leads the agglutination of RBCs when treated with Anti-A serum. The FACS analyses amply confirm the results obtained by the agglutinations tests, indicating that when agglutination was not observed the samples were completely converted. Indeed, samples with a partial agglutination have a rate of conversion above the 70%, highlighting the sensibility of agglutination tests performed.

5.5. Conclusion

In this work, three new GH109 have been characterized and show superior kinetic properties on 4NP-αGalNAc regarding the same family's enzymes already characterized. Furthermore, we have explored their capability to convert type A to type 0 RBCs. Preliminary results have shown that all the three enzymes can act on erythrocyte A antigens. On RBCs treated with NAg68 and NAg69, only in the presence of dextran agglutination was not observed. At the same time, NAg71, even in the absence of dextran, was able to act on RBCs type A and confirmed its superior power of conversion in the presence of dextran. These data showed that the three enzymes can act on erythrocyte to convert blood group A in the universal donor group 0. The NagA and α-NAGA GH109 enzymes from *E. meningosepticum,* previously used to convert RBCs, could fully convert blood group A at concentrations 10- and 3-folds higher (50 and 15 μg mL^{-1}, respectively), in the presence of dextran [107, 168]. Moreover, the enzymatic system converting more efficiently A-type needed at least 5 μg mL^{-1} of enzymes [169]. Instead, here, we showed that we did not observe agglutination in all samples treated with 2 μg mL^{-1} of NAg71, suggesting that this enzyme could represent an excellent biocatalyst capable to improve the biotechnological process of blood conversion. Further experiments are needed to reach a statistic data on more RBCs samples from different donors and to improve the enzymatic assays.

CHAPTER 6

6. Concluding Remarks

The impact of enzymes as biocatalysts in biotechnology has increased significantly in the development of industrial bioprocesses, indeed the enzyme global market is constantly expanding. In order to improve existing bioprocesses and develop new biotechnologies, the identification and characterization of novel enzymes with higher activity, stability, and specificity than the already used biocatalysts becomes indispensable. With this aim, my thesis work has described the identification and characterization of novel microbial glycoside hydrolases (GH) by using two different approaches (chapters 2 and 3), and more advanced studies on potential biotechnology applications of GHs (chapter 4 and 5).

The metagenomic study of two mud pools in Pisciarelli solfatara, described in chapter 2, offered the opportunity to discover novel Cazymes with new substrate specificities, reaction mechanisms, and optimal catalytic conditions. Among the other enzymes, this study allowed to identify and characterize the first archaeal member of the GH109 family (GH109_Pool2) showing characteristics that differ significantly from other members belonging to this family, such as the substrate specificity and the activation by exogenous NAD^+. Indeed, GH109_Pool2 is the first β-*N*-acetylglucosaminidase/β-glucosidase, expanding the activities repertoire of this CAZy family. This survey allowed the identification of 278 and 308 CAZymes sequences, from Pool 1 and 2, respectively, therefore it will be essential to foresee their enzymatic characterization to confirm the inferred activities and evaluate their exploitation in biotechnological application. The metagenomic approach in extreme environments is probably the most powerful tool to explore the biodiversity of other extreme environments in search for novel biocatalysts to be exploited in biotechnological applications operating in harsh conditions. However, considering the massive number of enzyme sequences identified so far, there are already many potential interesting enzymatic activities deposited in online databases and not yet explored. In this frame, the genome mining approach represents a valid alternative to discover novel biocatalysts. By using this approach, in the third chapter of the thesis, the identification and biochemical characterization of a new esterase from the thermophilic microorganism *G. thermodenitrificans* has reported . The biochemical characterization of the enzyme, namely EstGtA3, has revealed interesting exploitable properties for biotechnological applications. In fact, EstGtA3 was active at 60°C in a wide range of pH and showed an activating effect in *n*-hexane and other denaturing agents, making this enzyme suitable for resolution of racemic mixtures. This study demonstrated the effectiveness of *in silico*

analyses and bioinformatic strategies in the search for enzyme sequences already available in online databases, allowing to filter and preselect attractive targets to support experimental tests. In addition, it stresses the importance of enzyme characterization as the only tool that can provide and validate the properties of an enzyme as a biocatalyst.
In the framework of identification of new biocatalysts to be exploited in lignocellulose saccharification for second-generation biorefineries, the mechanism of action of three GHs from *S. solfataricus* in the hydrolysis of xyloglucan hemicellulose (XG) oligosaccharides was analyzed and reported in chapter 4. The three enzymes used in this study were already biochemically characterized in detail, however their combined action has never been tested on natural substrates. The study has revealed the efficient hydrolysis of the XG oligosaccharides by the cooperative action of the three enzymes with excellent thermal operational stability. In order to improve the study of lignocellulolytic enzymes, the work highlighted the importance of testing the activity of enzymes on natural substrates of industrial interests and of investigating their mechanism of action and synergism in the hydrolysis of such substrates.
In chapter 5 a more advanced biotechnological approach using GHs was analyzed. The work is in progress and is done in collaboration with Novozyme. In particular, three novel GH109 enzymes and their ability to convert blood group A in blood group 0, by removing the immunodominant antigen *N*-acetylgalactosamine from the surface of erythrocytes were characterized. Interestingly, one of the three enzymes, namely NAg71, was required in less amount than previously characterized enzymes capable of achieving this bioconversion, laying the foundation for improving biotechnology. This study could be an essential starting point to enable the large-scale development of this bioconversion to increase the availability of universal blood in blood centers and blood banks. Future experiments and further studies in the field will clarify the potential of this biotechnological application.
The thesis article strongly supports the need to identify new enzymatic activities by emphasizing the role of enzymatic characterization in searching for new biocatalysts. Therefore, to expand the set of commercially available biocatalysts, it is essential to combine different strategies to identify new enzyme sequences and to evaluate on a laboratory scale the ability of an enzyme to work on substrates and in chemical-physical conditions that could have relevance for biotechnological applications in the industrial field.

7. References

1. Robinson P K. Enzymes: principles and biotechnological applications. *Essays Biochem.* **2015**. 59:1–41.
2. Arbige M V, Shetty J K and Chotani G K. Industrial Enzymology: The Next Chapter. *Trends in Biotechnology*. **2019**. 37:12
3. Maester T C, Pereira M R, Malaman A M G, Borges J P, Pereira P A M and Eliana G. M. Lemos E G M L. Exploring Metagenomic Enzymes: A Novel Esterase Useful for Short-Chain Ester Synthesis. *Catalysts*. **2020**. 10:1100
4. Chapman J, Ismail A E and Dinu C Z. (2018). Industrial Applications of Enzymes: Recent Advances, Techniques, and Outlooks. *Catalysts*. **2018**. 8:238.
5. Sheldon R A. Biocatalysis and biomass conversion: enabling a circular economy. *Phil Trans R Soc A.* **2020**. 24;378(2176):20190274.
6. Wiltschi B, Cernava T, Dennig A, Casas M G, Geier M, Gruber S, Haberbauer M, Heidinger P, Acero E H, Kratzer R, Luley-Goedl C, Müller C A, Pitzer J, Ribitsch D, Sauer M, Schmölzer K, Schnitzhofer W, Sensen C W, Soh J, Steiner K, Winkler C K, Winkler M, Wriessnegger T. Enzymes revolutionize the bioproduction of value-added compounds: From enzyme discovery to special applications. *Biotechnology Advances*. **2020**. 40: 107520.
7. Liu X and Kokare C. Microbial Enzymes of Use in Industry. *Biotechnology of Microbial Enzymes*. **2017**. chapter 11; DOI: http://dx.doi.org/10.1016/B978-0-12-803725-6.00011-X
8. Min Jin M, Gai Y, Guo X, Hou Y and Zeng R. Properties and Applications of Extremozymes from Deep-Sea Extremophilic Microorganisms: A Mini Review. *Mar. Drugs*. **2019**. 17, 656
9. Marshall J R, Yao P, Montgomery S L, Finnigan J D, Thorpe T W, Palmer R B, Mangas-Sanchez J , Duncan R A M , Heath R S, Graham K M, Cook D J, Charnock S J and Turne N J. Screening and characterization of a diverse panel of metagenomic imine reductase for reductive amination. *Nature Chemistry*. **2021**. 13,140-148.
10. Baweja M, Nain L, Kawarabayasi Y and Pratyoosh S. Current Technological Improvements in Enzymes toward Their Biotechnological Applications. *Front Microbial*. **2016**. 16,7:965
11. Kries H, Blomberg R and Hilvert S. De novo enzymes by computational design. *Current Opinion in Chemical Biology*. **2013**. 17:221–228
12. Martinéz-Martinéz M, Bargiela R and Ferrer M. Metagenomic and the search for industrial enzymes. *Biotechnology of microbial enzymes*. **2017**. Chapter 7 https://doi.org/10.1016/B978-0-12-803725-6.00007-8
13. Margulies M, Egholm M, Altman W *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. **2005**. 437**,** 376–380.

14. Vincent A T, Derome N, Boyle B, Culley A I and Charette S J. Next-generation sequencing (NGS) in the microbiological world: How to make the most of your money. Journal of Microbiol Met. **2017**. 138:60-71.
15. Bentley D R, Balasubramanian S, Swerdlow H P, Smith G P, Milton J, Brown C G, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. **2008**; 456 (7218): 53-9.
16. Besser J, Carleton H A, Gerner-Smidt P, Lindsey R L and Trees E. Next-generation sequencing technologies and their application to the study and control of bacterial infection. Clinical microbiol and infection. **2018**. 24:335-341.
17. Ahmad T, Singh R S, Gupta G Sharma A and Kaur B. Metagenomics in the Search for Industrial Enzymes. *Advance in enzymes technology* Book. **2019**. Chapter 15: 419-451
18. Pimentel A C, Ematsu G C, Liberato M V et al. Biochemical and biophysical properties of a metagenome-derived GH5 endoglucanase displaying an unconventional domain architecture. *Int J Biol Macromol.* **2017**. 99:384–93
19. Zhao C, Chu Y, Li Y et al. High-throughput pyrosequencing used for the discovery of a novel cellulase from a thermophilic cellulose-degrading microbial consortium. *Biotechnol Lett*. **2017**. 39:123–31.
20. Cheng J, Huang S, Jiang H et al. Isolation and characterization of a non-specific endoglucanase from a metagenomic library of goat rumen. World J Microb Biot 2016;32:12.
21. Martin M, Biver S, Steels S et al. Identification and characterization of a halotolerant, cold-active marine endoβ-1,4-glucanase by using functional metagenomics of seaweed-associated microbiota. Appl Environ Microb 2014;80: 4958–67.
22. Lee CM, Lee YS, Seo SH et al. Screening and characterization of a novel cellulase gene from the gut microflora of Hermetia illucens using metagenomic library. J Microbiol Biotech 2014;24:1196–206
23. Schroder C, Elleuche S, Blank S ¨ et al. Characterization of a heatactive archaeal β-glucosidase from a hydrothermal spring metagenome. Enzyme Microb Technol 2014;57:48–54.
24. Li Y, Liu N, Yang H et al. Cloning and characterization of a new β-glucosidase from a metagenomic library of rumen of cattle feeding with Miscanthus sinensis. BMC Biotechnol 2014b;2:14– 85.
25. Sun MZ, Zheng HC, Meng LC et al. Direct cloning, expression of a thermostable xylanase gene from the metagenomic DNA of cow dung compost and enzymatic production of xylooligosaccharides from corncob. Biotechnol Lett 2015;37:1877–86.
26. Matsuzawa T, Kimura N, Suenaga H et al. Screening, identification, and characterization of α-xylosidase from a soil metagenome. J Biosci Bioeng 2016;122:393–9.
27. Lezyk M, Jers C, Kjaerulff L et al. Novel α-L-fucosidases from a soil metagenome for production of fucosydated human milk oligosaccharides. PLoS One 2016;11:e0147438.

28. Liu Z, Zhao C, Deng Y et al. Characterization of a thermostable recombinant β-galactosidase from a thermophilic anaerobic bacterial consortium YTY-70. Biotechnol Biotec Eq 2015b;29:547–54.
29. Gao W, Wu K, Chen L et al. A novel esterase from a marine mud metagenomics library for biocatalytic synthesis of shortchain flavoresters. Microb Cell Fact 2016b;15:41.
30. Zarafeta D, Moschidi D, Ladoukakis E et al. Metagenomic mining for thermostable esterolytic enzymes: uncovers a new family of bacterial esterases. Sci Rep 2016;6:38886.
31. De Santi C, Ambrosino L, Tedesco P et al. Identification and characterization of a novel salt-tolerant esterase from a Tibetan glacier metagenomic library. Biotechnol Prog 2015;31:890–9.
32. Preeti A, Hemalatha D, Rajendhran J et al. Cloning, expression and characterization of a lipase encoding gene from human oral metagenome. Indian J Microbiol 2014;54:284–92.
33. Brault G, Shareck F, Hurtubise Y et al. Short-chain flavour ester synthesis in organic media by an E. coli whole-cell biocatalyst expressing a newly characterized heterologous lipase. PLoS One 2014;9:e91872.
34. Thapa S, Li H, OHair J, Bhatti S, Chen F-C, Al Nasr K, Johnson T and Zhou S. Biochemical Characteristics of Microbial Enzymes and Their Significance from Industrial Perspectives. *Molecular Biotechnology*. **2019**. 61:579–601.
35. Hadadi N, MohammadiPeyhani H, Miskovic L, Seijo M and Hatzimanikatis V. Enzymes annotation for orphan and novel reactions using knowledge of substrate reactive sites. PNAS. **2019**. 116(15): 7298-7307.
36. Vorapreeda T, Thammarongtham C, Cheevadhanarak S and Laoteng K. Genome mining of fungal lipid-degrading enzymes for industrial applications. Microbiology. **2015**.161:1613–1626
37. Prayogo F A, Budiharjo A, Kusumaningrum H P, Wijanarka W, Suprihadi A and Nurhayati N. Metagenomic applications in exploration and development of novel enzymes from nature: a review. *J Genet Eng Biotechnol*. **2020**. 18:39
38. Kimura N. Novel Biological Resources Screened from Uncultured Bacteria by a Metagenomic Method. *Metagenomics Perspectives, Methods, and Applications* Book. **2018**. Chapter 14:273-288.
39. Niu S, Yang J, Mcdermaid A, Zhao J, Kang Y. Bioinformatics tools for quantitative and functional metagenome and metatranscriptome data analysis in microbes. *Brief Bioinform*. **2018**. 19:1415–1429
40. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W, MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. *Bioinformatics*. **2015**. 31(10):1674–1676.
41. Lin H-H and Liao Y-C. Accurate binning of metagenomic contigs via automated clustering sequences using information of genomic signatures and marker genes. *Sci Rep*. **2016**. 6:24175.

42. Wu Y-W, Simmons B A and Singer S W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*. **2016**. 32:605–7.
43. Miller I J, Rees E R, Ross J, Miller I, Baxa J, Lopera J, et al. Autometa: automated extraction of microbial genomes from individual shotgun metagenomes. *Nucleic Acids Res*. **2019**. 47:e57.
44. Metagenome analysis: a powerful tool for enzyme bioprospecting Madhavan A, Sindhu R, Parameswaran B, Sukumaran R K and Pandey A. *Appl Biochem Biotechnol*. **2017**. 183:636–651
45. Alves L F, Westmann C A, Lovate G L, Viana de Siqueira G M V, Borelli T C and Guazzaroni M-E. Metagenomic Approaches for Understanding New Concepts in Microbial Science. *Int J Genomics*. **2018**. doi: 10.1155/2018/2312987.
46. Johnson J S, Spakowicz D j, Hong B-Y, Petersen L M, Demkowicz P, Chen Lei, Leopold S R, Hanson B M, Agresta H O, Gerstein M, Sodergren E and Weinstock G M. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature comm*. **2019**. 10:5029.
47. Robinson S L, Piel J and Sunagawa S. A roadmap for metagenomic enzymes discovery. *Nat Prod Rep*. **2021**. DOI: 10.1039/d1np00006c.
48. Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol*. **2006**. 72:211–222.
49. Venter J C, Remington K, Heidelberg J F, Halpern A L, Rusch D, Eisen J A et al.. Environmental genome shotgun sequencing of the Sargasso Sea. *Science*. **2004**. 304:66–74.
50. Chen I-M A, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, et al. The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res*. **2021**. 49:D751–D763.
51. Reichart N L, Bowers R M, Woyke T and Hatzenpichler R. High potential biomass-degrading enzymes revealed by Hot Spring metagenomics. Front Microbiol. **2021**. 12:668238.
52. Kamble P and Vavilala S. Discovering novel enzymes from marine ecosystems: a metagenomic approach. *Botanica Marina*. **2018**. 61(2):161-175
53. Berini F, Casciello C, Marcone G L and Marinelli F. Metagenomics: novel enzymes from non-cultural microbes. *FEMS Microbiol Letters*. **2017**. 364:21
54. Ferrer M, Martinez-Martinez M, Bargiela R et al. Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb Biotechnol*. **2016**. 9:22– 34.
55. Ngara T R and Zhang H. Recent advances in function-based metagenomic screening. *Genom proteom bioinf*. **2018**. 16:405-415.
56. Gerlt J A. Tools and strategies for discovering novel enzymes and metabolic pathways. *Perspectives in Science*. **2016**. 9:24-32.

57. Vanacek P, Sebestova E, Babkova P, Bidmanova S, Daniel L, Dvorak P, Stepankova V, Chaloupkova R, Brezovsky J, Prokop Z and Damborsky J. Exploration of Enzyme Diversity by Integrating Bioinformatics with Expression Analysis and Biochemical Characterization. *ACS Catal*. **2018**. 8:2402-2412
58. Dorr M, Fibinger M P C, Last D, Schmidt S, Santos-Aberturas J, Bottcher D, Hummel A, Vickers C, Voss M and Bornscheuer U T. Fully Automatized High-Throughput Enzyme Library Screening Using a Robotic Platform. *Biotechnology and Bioengineering* 113,7
59. Vincentelli R, Cimino A, Geerlof A , Kubo A, Satou Y, Cambillau C. High-throughput protein expression screening and purification in Escherichia coli. *Methods*. **2011**. 55:65-72
60. Colin P-Y, Kintses B, Gielen F, Miton C M, Fischer G, Mohamed M F, Hyvo¨nen M, Morgavi D P, Janssen D B and Hollfelder F. Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nat Commun*. **2015**. https://doi.org/10.1038/ncomms10008.
61. Jacques P, Be´chet M, Bigan M, Caly D, Chataigne´G, Coutte F, Flahaut C, Heuson E, Lecle`re V, Lecouturier D, Phalip V, Ravallec R, Dhulster P and Froidevaux R. High-throughput strategies for the discovery and engineering of enzymes for biocatalysis. *Bioprocess Biosyst Eng*. **2017**. 40:161–180.
62. Sheludko Y V and Fessner W-D. Winning the numbers game in enzyme evolution – fast screening methods for improved biotechnology proteins. *Curr Opin Struct Biol*. **2020**. 63:123–133
63. Lobb B and Doxey A C. Novel function discovery through sequence and structural data mining. *Curr Opin Struct Biol*. **2016**. 38:53–61
64. Zaparucha A, De Berardinis V and Vaxelaire-Vergne C. Genome Mining for Enzyme Discovery. *Modern Biocatalysis Advances Towards Synthetic Biological Systems* Book. **2018**. Chapter 1: 1-27
65. Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science. **2011**. 331(6016):463-7.
66. Cruz L M, Trefflich S, Weiss V A, and Castro M A A. Protein Function Prediction. Methods Mol Biol. **2017**. *Funtional Genomics* Book 1654:55-75.
67. Finn R B, Bateman A, Clements J, Coggill P, Eberhardt R Y, Eddy S R, Heger A, Hetherington K, Holm L, Mistry J, Erik L. L. Sonnhammer E L L, Tate J and Punta M. Pfam: the protein families database. *Nucleic Acids Res*. **2014**. doi:10.1093/nar/gkt1223
68. Marchler-Bauer A, Derbyshire M K, Gonzales N R, Lu S, Chitsaz F, Geer L Y, Geer R C, He J, Gwadz M, Hurwitz D I, Lanczycki C J, Lu F, Marchler G H, Song J S, Thanki N, Wang Z, Yamashita R A, Zhang D, Zheng C and

Bryant S H. CDD: NCBI's conserved domain database. *Nucleic Acids Res*. **2015**. 43:D222–D226,.
69. Hawkins T and Kihara D. Function prediction of uncharacterized proteins. *J Bioinform Comput Biol*. **2007**. 5(1):1-30
70. Sillitoe I, Lewis T E, Cuff A, Das S, Ashford P, Dawson N L, Furnham N, Laskowski R A, Lee D, Lees J G, Lehtinen S, Studer R A, Thornton J and Orengo C A. CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res*. **2015**. 43:D376–D381.
71. Galperin M Y, Makarova K S, Wolf Y I and Koonin E V. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res*. **2015**. 43: D261–D269.
72. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. **2016**. 44:D457–D462
73. Lombard V, Golaconda Ramulu H, Drula E, Coutinho P M, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. **2014**. 42:D490–D495.
74. Fongaro G,Maia G A, Rogovski P, Cadamuro R D, Lopes J C, Moreira R S, Camargo A F, Scapini T, Stefanski F S, Bonatto C, Souza D S M, Stoco P H, Duarte R T D, da Cruz A C C , Wagner G and Treichel H. Extremophile Microbial Communities and Enzymes for Bioenergetic Application Based on Multi-Omics Tools. Current Genomics. **2020**. 21:240-252
75. D'Ischia M, Manini P, Moracci M, Saladino R, Ball V, Thissen H, Evans R, Puzzarini C and Barone V. Astrochemistry and Astrobiology: Materials Sciencein Wonderland? Int J Mol Sci. **2019**. 20:4079.
76. Merino N, Aronson H S, Bojanova D P, Feyhl-Buska J, Wong M L, Zhang S and Giovannelli D. Living at the Extremes: Extremophiles and the Limits of Life in a Planetary Context. Front Microbiol. **2019**. 10:780
77. Coker J. Extremophiles and biotechnology: Current uses and prospects. F1000 research. **2016**. 5:396
78. Salgaonkar B B, Sawant D T, Harinarayanan S and Braganca J M. Alpha-amylase production by extremely halophilic archaeon *Halococcus* strain GUVSC8. *Starch.* **2019**. 71:1800018
79. Walia A, Guleria S, Mehta P *et al.* Microbial xylanases and their industrial application in pulp and paper biobleaching: a review. *Biotech*. **2017**. 7**:**11.
80. Bano A, Chen X, Prasongsuk S *et al.* Purification and Characterization of Cellulase from Obligate Halophilic *Aspergillus flavus* (TISTR 3637) and Its Prospects for Bioethanol Production. *Appl Biochem Biotechnol.* **2019**. 189**:**1327–1337.
81. Schreck S D and Amy M. Grunden A M. Biotechnological application of halophilic lipases and thioesterases. *Appl Microbiol Biotechnol*. **2014**. 98:1011–1021.
82. Wani N, Shezad F, Ahanger F A, Malik S M, Sahay S and Jain K. Cold-active pectinase and their commercial application: A review. *Int. J. Sci. Res. in Biological Sciences.* 2018. 5(4):125-133.

83. Santiago M, Ramirez-Sarmiento C A, Zamora R A and Parra L P. Discovery, molecular mechanisms, and industrial application of cold-active enzymes. *Front Microbiol*. **2016**. 7:1408.
84. Al-Ghanayem A A and Joseph B. Current prospective in using cold-active enzymes as eco-friendly detergent additive. *Appl Microbiol Biotechnol.* **2020**. 104**:**2871–2882.
85. Mangiagalli M and Lotti M Cold-Active β-Galactosidases: Insight into Cold Adaptation Mechanisms and Biotechnological *Exploitation Mar Drugs* **2021**. 19:43.
86. Parashar D and Satyanarayana. An Insight into Ameliorating Production, Catalytic Efficiency, Thermostability and Starch Saccharification of Acid-Stable α-Amylases from Acidophiles. *Front Bioeng Biotechnol.* **2018**. 6:125
87. Msarah M J, Ibrahim I, Hamid A A and Aqma W S. Optimisation and production of alpha amylase from thermophilic *Bacillus* spp. and its application in food waste biodegradation. 2020. *Heliyon*. 6(6):e04183
88. Escuder-Rodríguez J J, DeCastro M E, Cerdán M E, Rodríguez-Belmonte E, Becerra M, González-Siso M I. Cellulases from Thermophiles Found by Metagenomics. *Microorganisms*. **2018**. 6(3):66.
89. Mathew, G.M., Madhavan, A., Arun, K.B. *et al.* Thermophilic Chitinases: Structural, Functional and Engineering Attributes for Industrial Applications. *Appl Biochem Biotechnol*. **2021**. 193:142–164.
90. Chadha B S, Kaur B, Basotra N, Tsang A and Pandey A. Thermostable xylanases from thermophilic fungi and bacteria: current perspective. *Bioresour Technol*. **2019**. 277:195-203.
91. Miguel-Ruano V, Rivera I, Rajkovic J, Knapik K, Torrado A, Otero J M, Beneventi E, Becerra M, Sánchez-Costa M, Hidalgo A, Berenguer J, González-Siso M I, Cruces J, Rúa M L, Hermoso JA. Biochemical and Structural Characterization of a novel thermophilic esterase EstD11 provide catalytic insights for the HSL family. *Comput Struct Biotechnol J*. **2021**. 10(19):1214-1232
92. Saqib S, Akram A, Halim SA, Tassaduq R. Sources of β-galactosidase and its applications in food industry. *Biotech*. **2017.** 7(1):79.
93. Elleuche S, Schröder C, Sahm K and Antranikian G. Extremozymes-biocatalysts with unique properties from extremophilic microorganisms. Curr Opin Biotechnol. **2014**. 29:116–123
94. Dumorné K, Córdova D C, Astorga-Eló M and Renganathan P. Extremozymes: a potential source for industrial applications. J. Microbiol. Biotechnol. **2017**. 27(4):649-659.
95. Atalaha J, Cáceres-Morenoa P, Espinaa G, Blamey J M. Thermophiles and the applications of their enzymes as new biocatalysts. Bioresource Technology. **2019**. 280:478–488
96. De Castro M-E, Rodríguez-Belmonte E and González-Siso M-I. Metagenomics of Thermophiles with a Focus on Discovery of Novel Thermozymes. Front Microbiol. **2016**. 7:1521.

97. Gomes D, Gama M and Domingues L. Determinants on an efficient cellulase recycling process for the production of bioethanol from recycled paper sludge under high solid loadings. *Biotechnol Biofuels*. **2018**. 11:111.
98. Cabrera M A and Blamey J M. Biotechnological applications of archaeal enzymes from extreme environments. Biol Res. **2018**. 51:37
99. Zarafeta D, Moschidi D, Ladoukakis E, Gavrilov S, Chrysina E D, Chatziioannou A, Kublanov I, Skretas G and Kolisis F N. Metagenomic mining for thermostable esterolytic enzymes uncovers a new family of bacterial esterases. Scientific Reports. **2016**. 6:38886
100. Zhu D, Adebisi W A, Ahmad F, Sethupathy S, Danso B and Sun J. Recent Development of Extremophilic Bacteria and Their Application in Biorefinery. Front Bioeng Biotechnol. **2020**. 8:483.
101. Escalapez J M, Bautista V, Torregrosa-Crespo J, Vegara A, Camacho M, Pire C, Bonete M J and Martinez-Espinosa R M. Extremophiles enzymes and biotechnology. *Extremophiles Enzyme and Biotechnology book*. 2018. DOI:10.1201/9781315154695-11.
102. Garron M-L and Henrissat B. The continuing expansion of CAZymes and their families. Curr Opin Chem Biol. **2019**. 53:82–87.
103. André I, Potocki-Véronèse G, Barbe S, Moulis C and Remaud-Siméon M. CAZyme discovery and design for sweet dreams. Curr Opin Chem Biol. **2014**. 19:17–24.
104. Stütz A E and Wrodnigg T M. Chapter 4 - Imino sugars and glycosyl hydrolases: Historical context, current aspects, emerging trends. Advances in Carbohydrate Chemistry and Biochemistry. **2011**. 66:187-298.
105. Koshland D E. Stereochemistry and the mechanism of enzymatic reaction. Biological Reviews. **1953**. 28:416-436.
106. Davies G and Henrissat B. Structures and mechanisms of glycosyl hydrolases. Structure. 1995. 3:853-9.
107. Liu Q P, Sulzenbacher G, Yuan H, Bennett E P, Pietz G, Saunders K, Spence J, Nudelman E, Levery S B, White T, Neveu J M, Lane W S, Bourne Y, Olsson M L, Henrissat B, and Clausen H. Bacterial glycosidases for the production of universal red blood cells. Nat Biotechnol. **2007**. 25, 454-64.
108. Teze D, Shuoker B, Chaberski E K, Kunstmann S, Fredslund F, Nielsen T S, Stender E G P, Peters G H J, Karlsson E N, Hededam D H and Hachern M A. The catalytic acid-base in GH109 resides in a conserved GGHGG loop and allows for comparable α-retaining and β-inverting activity in *N*-Acetylgalactosaminadase from *Akkermansia muciniphila*. *ACS Catal*. **2020**. 10,6:3809-3819
109. Aguilar A, Wohlgemuth R, Twardowski T. Perspectives on bioeconomy. *New Biotechnol*. **2018**. 40:181-184.
110. Stegman P, Londo M and Junginger M. The circular bioeconomy: Its elements and role in European bioeconomy. *Resour Conserv Recycl* **2020**. X 6:100029.

111. Becker S, Bouzdine-Chameeva and Jaegler A. The carbon neutrality principle: A case study in the French spirits sector. *J Clean Prod*. **2020**. 274:122739.
112. European Commission. A European Green Deal: Striving to be the first climate-neutral continent. **2019** [cited May 2021]; Available from: https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en
113. Botha J, Mizrachi E, Myburg A A and Cowan D A. Carbohydrate active enzyme domains from extreme thermophiles: components of a modular toolbox for lignocellulose degradation. *Extremophiles*. **2018**. 22: 1-12
114. Peng F, Peng P, Xu F and Sun R C. Fractional purification and bioconversion of hemicellulose. *Biotechnol Adv*. **2012**. 879-903.
115. Zoghlami A and Paes G. Lignocellulose biomass: Understanding recalcitrance and predicting hydrolysis. Front Chem. 2019. https://doi.org/10.3389/fchem.2019.00874
116. Mussatto S I, Dragone G, Guimarães P D R, Silva J P A, Carneiro L M, Roberto I C, Vicente A, Domingues L and Teixeira J A. Technological trends, global market, and challenges of bio-ethanol production. *Biotechnol Adv*. **2010**. 28: 817-30
117. Chen H and Fu X. Industrial technologies for bioethanol production from lignocellulosic biomass. *Renew Sust Energ Rev*. **2016**. 57: 468-78
118. Tan H, Corbin R and Ficher G B. Emerging Technologies for the Production of Renewable Liquid Transport Fuels from Biomass Sources Enriched in Plant Cell Walls. *Extremophiles*. **2018**. 22:1–12
119. Wang B, Wu S, Chang X, Chen J, Ma J, Wang P and Zhu G. Characterization of a novel hyper-thermostable and chlorpyrifos-hydrolyzing carboxylesterase EstC: A representative of the new esterase family XIX. Pestic Biochem Physiol. **2020**. 170:104704
120. Ramnath L, Sithole B and Govinden R. Classification of lipolytic enzymes and their biotechnological applications in the pulping industry. Can J Microbiol. **2017**. 63: 179–192.
121. Bornscheuer U T. Microbial carboxyl esterases: classification, properties and application in biocatalysis. FEMS Microbiology Reviews. **2002**. 26:73-81.
122. Arpigny J L and Jaeger K E. Bacterial lipolytic enzymes: classification and properties. Biochem J. **1999**. 343:177–183
123. Hitch C A T and Clavel T. A proposed update for the classification and description of bacterial lipolytic enzymes. PeerJ. **2019**. 7:e7249.
124. Samoylova Y V, Sorokina K N, Romanenko M V and Parmon V N. Cloning, expression and characterization of the esterase estUT1 from Ureibacillus thermosphaericus which belongs to a new lipase family XVIII. Extremophiles. **2018**. 22:271-285.
125. López-López O, Cerdán M-E and González-Siso M-I. Thermus thermophilus as a Source of Thermostable Lipolytic Enzymes. Microorganisms. **2015**. 3:792-808.

126. Giudicepietro F, Chiodini G, Caliro S, De Cesera W, Esposito A M, Galluzzo D, Lo Bascio D, Macedonio G, Orazi M, Ricciolino P and Vandemeulebrouck J. Insight into Campi Flegrei Caldera unrest through seismic tremor measurament at Pisciarelli Fumarolic field. *Geochm Geiphys Geosyst*. **2019**. 20:5544-5555.
127. Strazzulli A, Iacono R, Giglio R, Moracci M and Cobucci-Ponzano. Metagenomics of hyperthermophilic environments: biodiversity and biotechnology. *Microbial ecology of extreme environment* book. **2017**. Spinger, London (GBR).
128. Iacono R, Cobucci-Ponzano B, De Lise F, Curci N, Maurelli L, Moracci M and Strazzulli A. Spatial metagenomic of three geothermal sites in Pisciarelli hot spring on the biochemical resources of the microbial consortia. *Molecules*. **2020**. 25:4023.
129. Woese C R and Fox G E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *PNAS*. **1977**. 11:5088-5090.
130. Eme L, Spang A, Lombard J, Stairs C W and Ettema T J E. Archaea and the origin of eukaryotes. *Nat Rev Microbiol*. **2017**. 15:711-723.
131. Rampelotto P H. Extremophiles and extreme environment. *Life*. **2013**. 3:482-485.
132. Littlechild J A. Archaeal enzymes and applications in industrial biocatalysts. *Archaea.* **2015**. https://doi.org/10.1155/2015/147671.
133. Straub C T, Counts J A, Nguyen D M N, Wu C-H, Zeldes B N, Crosby J R, Conway J M, Otten J K, Lipscomb G L, Schut G J, Adams M W W and Kelly R M. Biotechnology of extremely thermophilic archaea. FEEMS. 2018. 42:543-578.
134. Suleiman M, Kruger A and Antranikian G. Biomass-degrading glycoside hydrolases of archaeal origin. *Biotechnol Biofuels*. **2020**. 13:153
135. Hyatt D, Chen G L, LoCascio P F, Land M L, Larimer F W and Hauser L J. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* **2010**. 11:119
136. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP et al. (2008) A core gut microbiome in obese and lean twins. Nature 457, 480.
137. Shi H, Huang Y, Zhang Y, Li W, Li X and Wang F. High-Level expression of a novel thermostable and mannose-tolerant beta-mannosidase from Thermotoga thermarum DSM 5069 in *Escherichia coli*. *Bmc Biotechnol*. **2013**. 13:83
138. Khan A, Singh P and Srivastava A. Synthesis, nature and utility of universal iron chelator—Siderophore: a review. *Microbiol Res*. **2018**. 212–213:103–111.
139. Tourova T P, Poltaraus A B, Novikova E V, Grigoryan A A, Ivanova A E, Lysenko A M, Petrunyaka V V, Osipov G A, Belyaev S S, Ivanov M V. Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus*

thermoleovorans, *Bacillus kaustophilus*, *Bacillus thermodenitrifcans* to *Geobacillus* as the new combinations *G. stearothermophilus*, G. th*. Int J Syst Evolut Microbiol*. **2001**. 51:433–446

140. Manachini PL, Mora D, Nicastro G, Parini C, Stackebrandt E, Pukall R, Fortina MG (2000) *Bacillus thermodenitrifcans* sp. nov., nom. rev. *Int J Syst Evolut Microbiol.* **2000**.50,3:1331–1337 doi:10.1099/00207713–50–3–1331.
141. Wissuwa J, Stokke R, Fedøy A-E, Lian K, Smalas A O and Steen I H. Isolation and complete genome sequence of the thermophilic Geobacillus sp. 12AMOR1 from an Arctic deep-sea hydrothermal vent site. *Stand Genomic Sci*. **2016**. 11:16.
142. Wang L, Tang Y, Wang S, Liu R L, Liu M Z, Zhang Y, Liang F L and Feng L. Isolation and characterization of a novel thermophilic *Bacillus* strain degrading long-chain n-alkanes. *Extremophiles*. **2006**. 10:347–356.
143. Thapa S, Mishra J, Arora N, Mishra P, Li H, Hair J, Bhatti S and Zhou S. Microbial cellulolytic enzymes: diversity and biotechnology with reference to lignocellulosic biomass degradation. *Rev Environ Sci Biotechnol*. **2020**. 19:621–648.
144. Cann I, Pereira G V, Abdel-Hamid A M, Kim H, Wefers D, Kanai T, Sato T, Bernardi R C, Atomi H and Mackie R I. Thermophilic Degradation of Hemicellulose, a Critical Feedstock in the Production of Bioenergy and Other Value-Added Products. *Appl Environ Microbiol*. **2020**. 86:e02296-19.
145. Sakai H D and Norio K N. *Saccharolobus caldissimus* gen. nov., sp. nov., a facultatively anaerobic iron-reducing hyperthermophilic archaeon isolated from an acidic terrestrial hot spring, and reclassification of *Sulfolobus solfataricus* as *Saccharolobus solfataricus* comb. nov. and *Sulfolobus shibatae* as *Saccharolobus shibatae* comb. nov. *Int J Syst Evol Microbiol.* **2018**. 68,1271–1278.
146. Zillig W, Stetter K O, Wunderl S, Schulz W, Priess H, Scholz I (1980) The *Sulfolobus*-"*Caldariella*" Group: taxonomy on the basis of the structure of DNA-dependent RNA polymerases*. Arch Microbiol.* **1980**. 125:259–269.
147. Zaparty M, Esser D, Gertig S, Haferkamp P, Kouril T, Manica A, Pham T K, Reimann J, Schreiber K, Sierocinski P, Teichmann D, van Wolferen M, von Jan M, Wieloch P, Albers S V, Driessen A J M, Klenk H-P, Schleper C, Schomburg D, van der Oost J, Wright P C and Siebers B. 'Hot standards'' for the thermoacidophilic archaeon *Sulfolobus solfataricus*. *Extremophiles*. **2010**. 14:119–142
148. Quhenberger J, Shen L, Albers S-V, Siebers B and Spadiut O. Sulfolobus – A Potential key organism in future biotechnology. *Front Microbiol*. **2017**. 8:2474.
149. Cannio R, Di Prizito N, Rossi M and Morana A. A xylan-degrading strain of *Sulfolobus solfataricus*: isolation and characterization of the xylanase activity. *Extremophiles*. **2004**. 8:117–124.

150. Kim M-S, Park J-T, Kim Y-W, Lee H-S, Nyawira R, Shin H-S, Park C-S, Yoo S-H, Kim Y-R, Moon T-W and Park K-H. Properties of a novel thermostable glucoamylase from the hyperthermophilic archaeon Sulfolobus solfataricus in relation to starch processing. *Appl Environ Microbiol.* **2004**. 70:3933–3940.
151. Haseltine C, Rolfsmeier M and Blum P. (1996). The glucose effect and regulation of alpha-amylase synthesis in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J Bacteriol*. **1996**. 178:945–950.
152. Moracci M, Nucci R, Febbraio F, Vaccaro C, Vespa N, La Cara F and Rossi M. Expression and extensive characterization of a β-glycosidase from the extreme thermoacidophilic archaeon *Sulfolobus solfataricus* in *Escherichia coli*: authenticity of the recombinant enzyme. *Enzyme Microb Technol*. **1995** 17, 992–997.
153. Moracci M, Cobucci-Ponzano B, Trincone A, Fusco S, De Rosa M, van Der Oost J, Sensen C W, Charlebois R L and Rossi M. Identification and molecular characterization of the first alpha -xylosidase from an archaeon. *J Biol Chem*. **2000**. 275:22082–22089.
154. Cobucci-Ponzano B, Trincone A, Giordano A, Rossi M and Moracci M. Identification of an archaeal α-L-fucosidase encoded by an interrupted gene. *J Biol Chem*. **2003**. 278(17),14622-31.
155. Tuomivaara S T, Yao K, O'Neill M A, York W S. Generation and structural validation of library of diverse xyloglucan-derived oligosaccharides, including an update on xyloglucan nomenclature. *Carbohydr Res*. **2015**. 402:56-66.
156. Fry C, York S, Albersheim P, Darvill A, Hayashi T, Joseleau J-P, Kato Y, Lorences P, Maclachlan A, McNeil M, Mort J, Reid G, Seitz U, Selvendran R, Voragen J and Whit A R. An unambiguous nomenclature for xyloglucan-derived oligosaccharides. *Physiol Plant*. **1993**. 89:1-3.
157. Mishra A and Malhotra A V. Tamarind xyloglucan: a polysaccharide with versatile application potential. *J Mater Chem.* **2009**. 19: 8528-36.
158. Benko Z, Siika-aho M, Viikari L and Réczey. Evaluation of the role of xyloglucanase in the enzymatic hydrolysis of lignocellulosic substrates. *Enzyme Microb Technol*. **2008**. 43,2:109-114
159. Zavyalova A V, Rykova S V, Luninab N A, Sushkovaa V I, Yarotskya S V and Berezina O V. Plant Polysaccharide Xyloglucan and Enzymes That Hydrolyze It. *Russ J Bioorg Chem.* **2019**. 45(7), 845–859.
160. Landsteiner K. Agglutination phenomena of normal human blood. *Wien Klin Wochenschr*. **2001**. 113:768–769.
161. Rahfeld P and Withers S G. Toward universal donor blood: Enzymatic conversion of A and B type to O type. *J Biol Chem*. **2020** 295,2:325–334.
162. Malinski T J and Singh H. Enzymatic conversion of RBCs by α-*N*-acetylgalactosaminidase from *Spirosoma Linguale*. Enzyme Res. **2019**. 2019:6972835. doi: 10.1155/2019/6972835.
163. Laxmi V. The Global Blood Industry. *BCC Res Rep*. **2017**. 1:1–239.

164. Roberts N, James S, Delaney M and Fitzmaurice. The global need and availability if blood products: a modelling study. *The lancet haematology*. **2019**. 6(12):E606-E615.
165. Lenny L L, Hurst R, Goldstein J, Benjamin L J and Jones R L. Single-unit transfusions of RBC enzymatically converted from group B to group O to A and O normal volunteers. Blood. **1991**. 77:1383–1388.
166. Lenny L, Hurst R, Goldstein J and Galbraith R. Transfusions to group O subjects of 2 units of red cells enzymatically converted from group B to group O. Transfusion **1994**. 34:209–214.
167. Lenny L, Hurst, R, Zhu A, Goldstein J and Galbraith R. (1995) Multiple-unit and second transfusions of red cells enzymatically converted from group B to group O: report on the end of phase 1 trials. *Transfusion*. **1995**. 35:899–902.
168. Yu C Y , Xu H, Wang L S, Zhang J G and Zhang Y P. Human RBCs blood group conversion from A to O using a novel α-*N*-acetylgalactosaminidase of high specific activity. *Chi Sci Bull*. **2008**. 53,13:2008-2016.
169. Rahfeld P, Sim L, Moon H, Constantinescu I, Morgan-Lang C, Hallam S J, KizhakkedathuJ N and Withers S G. An enzymatic pathway in the human gut microbiome that converts A to universal O type blood. *Nat microbiol*. **2019**. 4:1475-1485.
170. Chapanian R, Kwan D H, Constantinescu I, Shaikh F A, Rossi N A A, Withers S G and Kizhakkedathu J N. Enhancement of biological reactions on cell surfaces via macromolecular crowding. *Nat Commun*. **2016**. 5:4683.

8. APPENDIX

8.1. Pubblications

- Iacono R, Strazzulli A, Maurelli L, **Curci N**, Casillo A, Corsaro M, Moracci M and Cobucci-Ponzano B. Identification and characterization of a GlcNAc de-N-acetylase from hyperthermophilic archaeon *Sulfolobus solfataricus*. *Appl Environ Microbiol*. **2018**. 85(2):e01879-18.

- **Curci N**, Strazzulli A, De Lise F, Iacono R, Maurelli L, Dal Piaz F, Cobucci-Ponzano B and Moracci M. Identification of a novel esterase from thermophilic bacterium *Geobacillus thermodenitrificans* NG80-2. *Extremophiles*. **2019**. 23:407-419.

- Strazzulli A, Cobucci-Ponzano B, Iacono R, Giglio R, Maurelli L, **Curci N**, Schiano-di-Cola C, Santangelo A, Contursi P, Lombard V, Henrissat B, Lauro F M, Fontes C M G A and Moracci M. Discovery of hyperstable carbohydrate-active enzymes through metagenomics of extreme environments. *The FEBS Journal* **2019** *287*,1116–1137.

- Iacono R, Cobucci-Ponzano B, De Lise F, **Curci N**, Maurelli L, Moracci M, Strazzulli A. Spatial metagenomics of three geothermal sites in Pisciarelli hot springs focusing on the biochemical resources of the microbial consortia. *Molecules*. **2020**. 25:4023.

- **Curci N**, Strazzulli A, Iacono R, De Lise F, Maurelli L, Di Fenza M, Cobucci-Ponzano B and Moracci M. Xyloglucan oligosaccharides hydrolysis by exo-acting glycoside hydrolases from hyperthermophilic microorganism *Saccharolobus solfataricus*. *Int J Mol Sci*. **2021**. 22:3325.

- De Lise F, Iacono R, Strazzulli A, Giglio R, **Curci N**, Maurelli L, Avino R, Carandente A, Caliro S, Tortora A, Lorenzini F, Di Donato P, Moracci M and Cobucci-Ponzano B. Transcript regulation of recoded archaeal α-*L*-Fucosidase in vivo. *Molecules*. **2021**. 26:1861.

8.2. Communication

- **Curci N**, Iacono R, Strazzulli A, Maurelli L, Cobucci-Ponzano B and Moracci M. Identification and characterization of a novel esterase from *Geobacillus thermodenitrificans*. 12th International Congress of Extremophiles 2018.

- **Curci N**, Iacono R, Strazzulli A, Maurelli L, Cobucci-Ponzano B and Moracci M. Identification and characterization of carbohydrate active enzymes for xyloglucan degradation. 12th International Congress of Extremophiles 2018.

- **Curci N**, Strazzulli A, Iacono R, De Lise F, Maurelli L, Cobucci-Ponzano B and Moracci M. Identification and characterization of the first archaeal Glycoside Hydrolase 109 from metagenomic dataset of Pisciarelli Solfatara. Workshop BIO/10 Campania 2019.

- **Curci N**, Strazzulli A, Iacono R, De Lise F, Maurelli L, Cobucci-Ponzano B and Moracci M. Novel esterase from the thermophilic bacterium *Geobacillus thermodenitrificans*. Workshop BIO/10 Campania 2019.

- **Curci N**, Iacono R, Strazzulli A, Maurelli L, De Lise F, Cobucci-Ponzano B and Moracci M. Xyloglucan hemicellulose degradation by Carbohydrate Active Enzymes from hyperthermophilic archaeon *Saccharolobus solfataricus*. CBM13 - Carbohydrate Bioengineering Meeting - Toulouse 2019.

- **Curci N**, Strazzulli A, Iacono R, De Lise F, Maurelli L, Cobucci-Ponzano B and Moracci M. Metagenomic of extreme environments as a tool for the discovery of new hyperstable enzymes. 1st Italian Space Agency workshop on Astrobiology 2020.

8.3. Period Abroad

From June 3rd to December 3rd, 2019; at Institute of Architecture et Function de Macromolécules Biologiques (AFMB) of the Centre National de la Recherche Scientifique (CNRS), University of Marseille. Under the supervision of Dr Gerlind Sulzenbacher.

8.4. Research Articles

8.4.1. **Discovery of hyperstable carbohydrate-active enzymes through metagenomics of extreme environments.** Strazzulli A, Cobucci-Ponzano B, Iacono R, Giglio R, Maurelli L, Curci N, Schiano-di-Cola C, Santangelo A, Contursi P, Lombard V, Henrissat B, Lauro F M, Fontes C M G A and Moracci M. *The FEBS Journal* **2019** *287*,1116–1137.

8.4.2. **Identification and characterization of a GlcNAc de-N-acetylase from hyperthermophilic archaeon Sulfolobus solfataricus.** Iacono R, Strazzulli A, Maurelli L, Curci N, Casillo A, Corsaro M, Moracci M and Cobucci-Ponzano B. *Appl Environ Microbiol*. 2018. 85(2):e01879-18.

8.4.3. **Spatial metagenomics of three geothermal sites in Pisciarelli hot springs focusing on the biochemical resources of the microbial consortia.** Iacono R, Cobucci-Ponzano B, De Lise F, Curci N, Maurelli L, Moracci M, Strazzulli A. *Molecules*. 2020. 25:4023

8.4.4. **Transcript regulation of recoded archaeal α-L-Fucosidase in vivo.** De Lise F, Iacono R, Strazzulli A, Giglio R, Curci N, Maurelli L, Avino R, Carandente A, Caliro S, Tortora A, Lorenzini F, Di Donato P, Moracci M and Cobucci-Ponzano B. *Molecules*. 2021. 26:1861.

The FEBS Journal

FEBS PRESS science publishing by scientists

Discovery of hyperstable carbohydrate-active enzymes through metagenomics of extreme environments

Andrea Strazzulli[1,2], Beatrice Cobucci-Ponzano[3], Roberta Iacono[1,3], Rosa Giglio[3], Luisa Maurelli[3], Nicola Curci[1,3], Corinna Schiano-di-Cola[3], Annalisa Santangelo[1], Patrizia Contursi[1,2], Vincent Lombard[4], Bernard Henrissat[4,5], Federico M. Lauro[6,7], Carlos M.G.A. Fontes[8,9] and Marco Moracci[1,2,3]

1 Department of Biology, University of Naples "Federico II", Complesso Universitario di Monte S. Angelo, Naples, Italy
2 Task Force on Microbiome Studies, University of Naples Federico II, Italy
3 Institute of Biosciences and BioResources – National Research Council of Italy, Naples, Italy
4 Centre National de la Recherche Scientifique, INRA, AFMB, USC 1408, Aix Marseille Univ, France
5 Department Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia
6 Asian School of the Environment, Nanyang Technological University, Singapore City, Singapore
7 Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore City, Singapore
8 NZYTech LDA, Estrada Do Paco Do Lumiar, Lisbon, Portugal
9 CIISA - Faculdade de Medicina Veterinária, Universidade de Lisboa, Portugal

Keywords
archaea; CAZymes; extremozyme; lignocellulose; microbiome

Correspondence
A. Strazzulli, Department of Biology, University of Naples "Federico II", Complesso Universitario di Monte S. Angelo, Via Cupa Nuova Cinthia 21, 80126 Naples, Italy
Tel: +39-081-679054
E-mail: andrea.strazzulli@unina.it

Present address
GSK VACCINES S.R.L, Siena, Italy
Department of Science and Environment, Roskilde University, Denmark

(Received 24 May 2019, revised 23 July 2019, accepted 1 October 2019)

doi:10.1111/febs.15080

The enzymes from hyperthermophilic microorganisms populating volcanic sites represent interesting cases of protein adaptation and biotransformations under conditions where conventional enzymes quickly denature. The difficulties in cultivating extremophiles severely limit access to this class of biocatalysts. To circumvent this problem, we embarked on the exploration of the biodiversity of the solfatara Pisciarelli, Agnano (Naples, Italy), to discover hyperthermophilic carbohydrate-active enzymes (*CAZymes*) and to characterize the entire set of such enzymes in this environment (*CAZome*). Here, we report the results of the metagenomic analysis of two mud/water pools that greatly differ in both temperature and pH (T = 85 °C and pH 5.5; T = 92 °C and pH 1.5, for Pool1 and Pool2, respectively). DNA deep sequencing and following *in silico* analysis led to 14 934 and 17 652 complete ORFs in Pool1 and Pool2, respectively. They exclusively belonged to archaeal cells and viruses with great *genera* variance within the *phylum* Crenarchaeota, which reflected the difference in temperature and pH of the two Pools. Surprisingly, 30% and 62% of all of the reads obtained from Pool1 and 2, respectively, had no match in nucleotide databanks. Genes associated with carbohydrate metabolism were 15% and 16% of the total in the two Pools, with 278 and 308 putative CAZymes in Pool1 and 2, corresponding to ~ 2.0% of all ORFs. Biochemical characterization of two CAZymes of a previously unknown archaeon revealed a novel subfamily GH5_19 β-mannanase/β-1,3-glucanase whose hemicellulose specificity correlates with the vegetation surrounding

Abbreviations
4NP-α-Gal, 4-Nitrophenyl α-D-galactopyranoside; 4NP-β-Gal, 4-Nitrophenyl β- D-galactopyranoside; 4NP-β-GalNAc, 4-Nitrophenyl *N*-acetyl-β-D-galactosaminide; 4NP-α-Glc, 4-Nitrophenyl α-D-glucopyranoside; 4NP-α-GlcNAc, 4-Nitrophenyl *N*-acetyl-α-D-glucosaminide; 4NP-α-Man, 4-Nitrophenyl α-D-mannopyranoside; 4NP-β-Glc, 4-Nitrophenyl β- D-glucopyranoside; 4NP-β-GlcNAc, 4-Nitrophenyl *N*-acetyl-β- D-glucosaminide; 4NP-β-Man, 4-Nitrophenyl β- D-mannopyranoside; AA, Auxiliary activity; CAZy, carbohydrate-active enzyme; CBM, carbohydrate-binding module; CE, carbohydrate esterase; COG, clusters of orthologous groups; GH, Glycoside hydrolase; Glc2, Laminaribiose; Glc3, Laminaritriose; Glc4, Laminaritetraose; Glc5, Laminaripentaose; GT, glycosyltransferase; KEGG, Kyoto Encyclopedia of Genes and Genome; MFS, major facilitator superfamily; NR, NCBI 'not-redundant' database; NT, NCBI nucleotide database; ORF, Open reading frame.

the sampling site, and a novel NAD^+-dependent GH109 with a previously unreported β-*N*-acetylglucosaminide/β-glucoside specificity.

Databases

The sequencing reads are available in the NCBI Sequence Read Archive (SRA) database under the accession numbers SRR7545549 (Pool1) and SRR7545550 (Pool2). The sequences of GH5_Pool2 and GH109_Pool2 are available in GenBank database under the accession numbers MK869723 and MK86972, respectively. The environmental data relative to Pool1 and Pool2 (NCBI BioProject PRJNA481947) are available in the Biosamples database under the accession numbers SAMN09692669 (Pool1) and SAMN09692670 (Pool2).

Introduction

Extremophilic microorganisms and their corresponding enzymes (extremozymes) have generated a great interest in the field of biocatalysis and biotransformation for their ability to function under conditions in which enzymes of mesophilic origin would quickly denature. Besides being thermostable, extremozymes show remarkable resistance to extreme pH, detergents, and high salts and organic solvent concentrations [1]. Thus, extremozymes are ideal tools for industrial applications where harsh chemical and physical conditions are encountered, such as the pulp and paper industry or starch and lignocellulose conversions [2–4]. Global climate change and the finite resources of fossil liquid fuels increase the demand for sustainable energy and chemical precursors for plastic materials. Among the different type of natural sources, lignocellulose is considered one of the most competitive alternatives to fossil fuels. In fact, nonedible lignocellulosic feedstocks (energy crops) constitute a virtually unlimited and convenient alternative to starchy materials (food crops) [5]. This realization led to second-generation biorefineries which, producing biofuel and bioplastic precursors from lignocellulose biomass rather than starch, may represent a solution to the food *vs* fuel problem [6–8].

The main components of lignocellulose are cellulose (35–50%), hemicellulose (20–35%), and lignin (10–25%), with small percentages of pectin depending on species, growth, and tissue [9]. Cellulose and hemicelluloses (xyloglucan, xylan, mannans, glucomannans, and β-(1,3-1,4)-glucans) are made of equatorial β-(1,4)-linked backbone structure while pectins (galactans, arabinans, and arabinogalactans) contain axial and equatorial backbone structures. In second-generation biorefineries, C6 and C5 sugars are produced by chemico-physical pretreatment (steam explosion and pretreatment with acid or alkali) of the lignocellulose biomass and then fermented to ethanol or bioplastic precursors [9]. Because of the recalcitrant nature of plant biomass, the complete hydrolysis of the lignocellulose is one of the main bottlenecks to fully realize this industrial process. Thus, the discovery of novel and ingenious enzymatic mixtures is essential to fully realize the process and to improve final yields. Presently, thermostable and thermoactive carbohydrate-active enzymes (CAZymes) are recognized as the main players in second-generation biorefineries [10,11]. In fact, their resistance to extreme conditions makes them ideal candidates for exploitation during the pretreatment step, by increasing efficiency, reducing the byproducts, and improving the final costs of the entire process.

The interest in extremophiles has boosted metagenomics studies of extreme environments [12–16], and, more recently, remarkable examples of the discovery of novel extremozymes for industrial applications from sequence-based screenings have been reported [17–19].

CAZymes include glycoside hydrolases (GHs), carbohydrate esterases (CE), polysaccharide lyases, and auxiliary activities and are classified in the CAZy database (www.cazy.org) including also carbohydrate-binding modules (CBM). [20]. In the framework of the discovery of new CAZymes from extremophiles for biocatalysis and biotransformations, we embarked in the exploration of the diversity of the solfataric field of Pisciarelli, Agnano (Naples, Italy), which witnessed pioneering works for the isolation of hyperthermophilic Archaea [21–24].

Here, we report the results of the metagenomic analysis of two mud/water pools that, although very close to each other, display different temperature and pH values (T = 85 °C and pH 5.5; T = 92 °C and pH 1.5, for Pool1 and Pool2, respectively). The whole sequence-based analysis confirmed the presence of archaeal strains previously isolated in this hydrothermal vent but showed unexpected ratios among them, and at the same time, the presence of many sequences which have no match with any sequence available in the NCBI database, revealing not only the existence of

a great untapped pool of genes but also a rich biochemical biodiversity even in these frequently sampled hydrothermal vents.

Moreover, the bioinformatic analysis of the CAZymes present in solfatara Pisciarelli, supported by the analysis of the vegetation surrounding the sampled site, allowed outlining the polysaccharide-degrading potential of this microbial community. Besides, the *in silico* analysis allowed us to identify two novel members of the GH5 and GH109 families that were produced and characterized in detail.

Results and Discussion

Community composition

Study site and sampling

The objective of this study was to use a metagenomic approach to analyze the hyperthermophilic (optimally growing at $T > 80$ °C) microbial communities in the solfatara Pisciarelli, Naples Italy (40°49′45.0768″N, 14°8′49.3512E), and to identify their CAZymes portfolio (Fig. 1). Therefore, we identified two hydrothermal mud/water pools, called Pool1 and Pool2, respectively, which were located at a distance of about 4 m from each other. Samples from Pool1 ($T = 85$ °C pH 5.5), which were taken from the surface of the pool, consisted mainly of hot muddy sediments, gravel, and water. Pool2 was sampled by scraping the side of the pool submerged by a clear mixture of mud/water ($T = 92$ °C pH 1.5) with a spoon and consisted mainly of gravel. Total DNA was extracted from the two samples and deep sequenced as described in the Materials and methods.

Whole metagenomic sequencing

The microbial composition of Pool1 and Pool2 samples was analyzed by performing a metagenomic sequence-based approach by Illumina (San Diego, CA, USA) HiSeq sequencing, followed by community diversity analysis based on assigning reads to known microorganisms of the NCBI nucleotide and protein database (*NT* and *NR*, respectively). The assembly rarefaction analysis, performed by Nonpareil 3 [25], revealed that both samples were sequenced at sufficient depth (Fig. S1) to capture the taxonomic diversity of each pool.

More than 30% and almost 50% of all of the obtained reads from Pool1 found no match in the known *NT* and *NR* NCBI databases (Fig. 2), respectively, while the phylogenetic analysis revealed that all others belonged to Archaea (67%) and only a small percentage to Bacteria (0.11%) and archaeal viruses (0.17%) (Fig. S2A). The Archaea domain is almost exclusively represented by Crenarchaeota (67%) (Fig. S2B), and, within this phylum, Sulfolobales (42%) are followed by Thermoproteales (25%) (Fig. S2D). The few bacterial families included Alphaproteobacteria, Enterobacteriaceae, and Clostridiales, (Fig. S2D). In particular, the representatives of *Acidianus hospitalis* (40%) and *Pyrobaculum arsenaticum* (20%), from Sulfolobales and Thermoproteales, respectively, were highly abundant among known microorganisms, followed by *Pyrobaculum oguniense* (5%) and *Sulfolobus solfataricus* (1%) (Fig. 2).

The majority of all of the Pool2 reads obtained from the Illumina sequencing had no match in both known NCBI nucleotide (62%) and protein (68%) databases

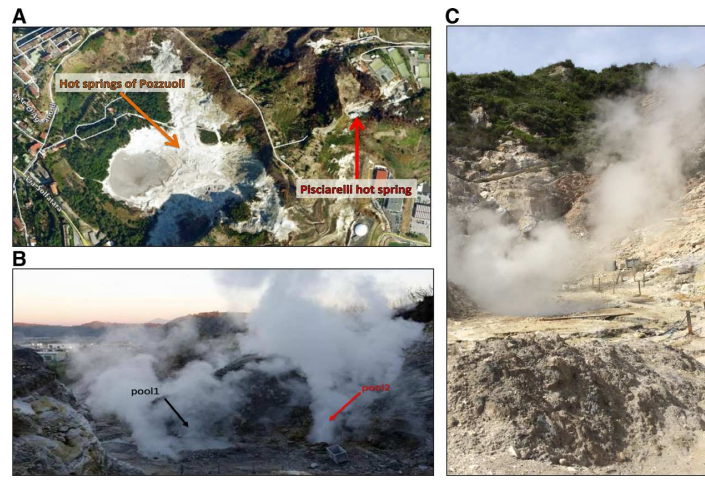


Fig. 1. The sampling site of solfatara Pisciarelli. The solfatara Pisciarelli is part of a wide volcanic site (Phlegrean Fields), which include also the solfatara of Pozzuoli (A). Pool1 and 2 are two hydrothermal vents showing strong emission of gas and vapor distant only 4 m (B). The crater of Pisciarelli solfatara is surrounded by rich local vegetation that is a potential source of energy for the microbial population (C).

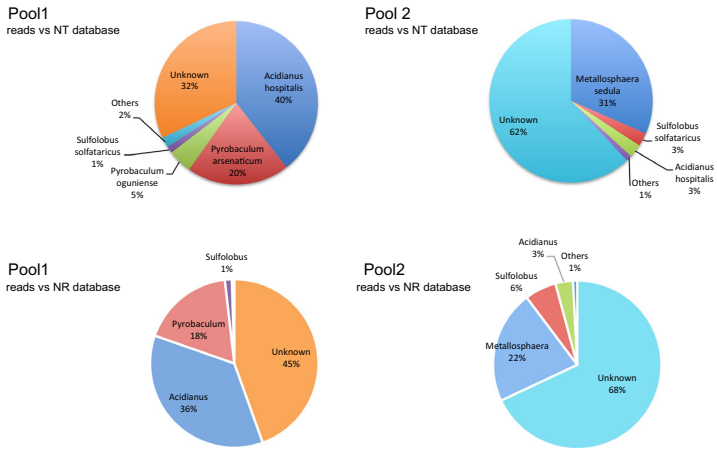


Fig. 2. Read composition of the metagenome of solfatara Pisciarelli. The percentage of the reads found in Pool1 and Pool2 classified at the genus level are reported. NR, NCBI 'not-redundant' database; NT, NCBI nucleotide database.

(Fig. 2). Remarkably, 49% of all the reads produced contigs, indicating the presence of large portions of sequences of unknown origin. Almost all the remaining reads, which could be assigned to known organisms, exclusively belonged to archaeal cells and viruses (37% and 0.36%, respectively) (Fig. S2). As in Pool1, Crenarchaeota dominated with the Sulfolobales order representing almost the totality of the microorganisms present. It is worth mentioning that *Metallosphaera sedula* (31% of all available reads), which was reported to grow at temperatures not higher than 85 °C [23] was, by far, the dominating species in Pool2 (92 °C), followed by *S. solfataricus* (3%), and *A. hospitalis* (3%) (Fig. 2). Other Archaea, including Thermoproteales, were present in traces (0.07%) (Fig. S2).

To estimate the biodiversity of the two samples analyzed by the whole metagenome sequencing, the Shannon–Weaver (H', indicating the species richness) and Simpson's reciprocal (1/D, indicating the even distribution of the microorganisms) diversity indexes were calculated on random subsets for each sample. The values (H' = 3.814 and 1/D = 12.580, H' = 3.578 and 1/D = 10.783 for Pool1 and Pool2, respectively), indicating lower diversity in Pool2 than Pool1, were not surprising, considering the extreme temperatures and pH of Pool2, which selected only organisms thriving at T 92 °C and pH 1.5.

The viral population of the two pools was searched through the analysis of the whole reads *vs* nucleotide database. Pool1 was exclusively populated by *Bicaudiviridae* (*Acidianus two-tailed virus* ATV) and *Lipothrixiviridae* (*Acidianus filamentous virus 7*-AFV7) members (92% and 8% of all the viral reads, respectively) [26], whereas Pool2 showed higher viral heterogeneity. Indeed, besides sharing with Pool 1 the presence of the same viruses, also hosted an *Ampullaviridae* representative, that is *Acidianus bottle-shaped virus* (13%), as well as other *Lipothrixiviridae* members such as AFV3 (2%), AFV6 (2%), and AFV8 (1%) [27].

To evaluate the coverage (cov) of the viral genomes, all the previously obtained contigs were aligned by using Bowtie2 [28], against all the viral genomes available in NCBI, revealing that in both samples the most covered viral genomes are ATV (cov. 21% and 34% in Pool1 and Pool2, respectively) (Table S1 and S2). Surprisingly, the latter analysis revealed in Pool1 also the presence of sequences related to *Fuselloviridae* members [29,30], that is, *Sulfolobus spindle-shaped virus* 2, SSV2 (cov. 15%), *Sulfolobus spindle-shaped virus* 5, SSV5 (cov. 5%), and *Sulfolobus monocaudavirus SMV1* (cov. 2%) that were not observed in the screening of the reads due to their low abundance of 0.01% of the whole assigned reads (MIN SUPPORT value in MEGAN6, see Material and Methods). In addition, *SMV1* was identified also in Pool2 sample with an overall coverage of 8% (Table S1 and S2). Overall the diversity of the viruses found in Pool 1 and Pool 2 is consistent with that of their potential hosts. Nevertheless, *Fuselloviridae* representatives have been only found in Pool1, despite the presence of their *S. solfataricus* host in both Pool1 and Pool2. No systematic studies on the stability of SSV virus particles to extreme temperature and pH are available in the literature; however, it is possible that SSV virus particles do not withstand the combination of the harsh chemico-physical conditions of Pool2, that is, low pH and high temperature that are too extreme compared to those of the environments in which *Sulfolobus* genera usually thrive. By contrast, ATV virus particles are present in both Pool 1 and Pool 2; this is not surprising since this

archaeal virus was discovered in an extremely acidic hot spring (85–93 °C; pH 1.5) and is one of the few archaeal viruses engaging a lytic lifestyle. Therefore, ATV virus particles are adapted to survive even in hostile environments [31]. The absence of viral sequences related to viruses infecting *Metallosphaera* and *Pyrobaculum* could be traced back to the establishment of weak or less stable host–virus interactions as previously described for these genera [32,33].

On the whole, this metagenomic sequencing demonstrates that the two pool samples display a striking difference in microbial composition, presumably mirroring their different temperature and pH conditions; thus, we performed a functional analysis of the available metagenomic data in order to get insight into the metabolic biodiversity of the microbial population.

Unassigned reads

Most of the reads from Pool1 and Pool2 (84% and 87%, respectively) could be successfully aligned to the obtained contigs, but one-third and two-thirds of all reads in Pool1 and 2, respectively, did not match any known sequence.

To identify possible individual genomes and genomes of related organisms, the contigs (≥ 1000 bp) were analyzed by MyCC [34] identifying 17 and 11 clusters in Pool1 and Pool2, respectively (Fig. S3). The subsequent analysis by CheckM [35] revealed that, among them, clusters 4, 6, and 8 from Pool1 and cluster 4 from Pool2 show near completeness (> 90%) with low contamination (< 5%) (Fig. 3, Table S3). This analysis allowed us to identify the taxonomic composition of these clusters.

The three clusters of Pool1 were taxonomically classified at the genera level as *Pyrobaculum* (cluster 4), *Acidianus* (cluster 6), and *Desulfurococcus* (cluster 8) while cluster 4 from Pool2 was classified as *M. sedula* with 97.62% of completeness, 1.79% of contamination, and 100% of strain heterogeneity (Table S3). These data strongly suggest that the low contamination in Pool2 cluster 4 could be ascribed to different strains of the same microorganism present in Pool2 sample.

In addition, although the completeness of Pool2 cluster 9 was moderate (55.36%), it presents very low contamination (0.6%) with no strain heterogeneity. It is worth noting that > 73% of the Pool2 reads showing no match in *NT* (12 116 649) are aligned to cluster 9 (Fig. S4) and that the analysis against *NR* database of its amino acid sequences showed identity between 26% and 100% with several members of the *Sulfolobaceae* family, namely *Sulfolobus*, *Sulfodiicoccus*, *Metallosphaera*, *Acidianus*, and *Candidatus aramenus*. Taken together, these results strongly suggest the possible presence of a novel microorganism not yet identified belonging to the *Sulfolobaceae* family.

By contrast, 43.64% of the Pool1 reads showing no match in *NT* (4 642 301) were aligned to cluster 6 (Fig. S4). The analysis of the predicted amino acid sequences of this cluster (1415) showed that they can be assigned to the genus *Acidianus* (98%) and more specifically to *A. hospitalis* (80%) with identity between 33% and 100%, indicating possible horizontal gene transfer.

To more broadly evaluate the whole biodiversity of the microbial population, the reads showing no match *vs NT* database were extracted and reassembled separately to obtain contigs. By following this approach, 44 550 and 73 287 contigs (cutoff > 100 bp) were obtained in Pool1 and Pool2, respectively. The gene prediction analysis revealed the presence of 43 118 putative ORFs in Pool1 and 84 140 in Pool2. Then, the ORFs identified (cutoff > 14 aa) were analyzed against the *NR* database to provide a taxonomical identification based on the amino acid sequence. This allowed the assignment of 37 667 and 76 319 sequences in Pool1 and Pool2, respectively, with a broad percentage of identity between 20% and 100% and an average

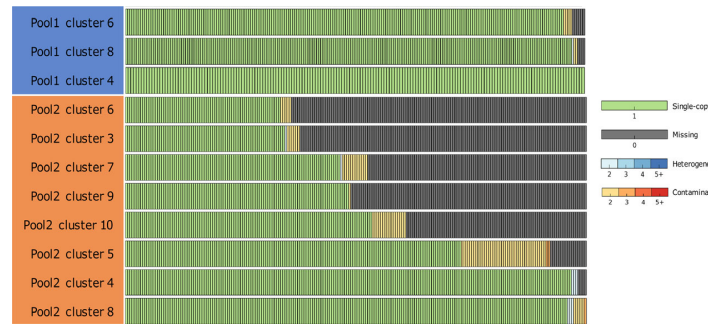


Fig. 3. Graphical representation of the clusters validated by CheckM with completeness > 20%. Bars in green represent markers identified exactly once. Markers identified multiple times in the genome are represented by shades of blue or red depending on the amino acid identity between pairs of multicopy genes and the total number of copies present (2−5+) (Table S3). Bars in gray represent missing markers.

Blast+/BlastP score of 119.2 (Pool1) and 172.1 (Pool2) (Table S4, Fig. S5, and S6).

The ORF-based taxonomic assignments at the genus level revealed that most of the sequences in Pool1 were related to sequences from *Acidianus* (45%) followed by *Sulfolobus* (20%) and *Desulfurococcus* (17%). By contrast, in Pool2 the ORFs identified pointed to sequences belonging to *Sulfolobus* (47%), *Acidianus* (28%), and *Metallosphaera* (11%) (Fig. S7). In particular, amino acid sequences related to *Desulfurococcus* and *Staphylothermus* could be only found in Pool1 (1.4%) indicating the presence of microorganisms belonging to the Desulfurococcaceae family. It is worth noting that, according to the whole-reads analysis, only 0.01% of the reads were assigned to the Desulfurococcaceae (Fig. S2E); presumably, novel members of this family, not yet identified, populate Pool1.

Functional annotation

The number of complete open reading frames (ORFs) identified in the whole assembly was 14 934 and 17 652 in Pool1 and Pool2, respectively. They were subjected to BLAST+/BlastP similarity searches and annotated with the Clusters of Orthologous Groups and the Kyoto Encyclopedia of Genes and Genomes database (COGs and KEGG, respectively) for a phylogenetical and functional classification (Fig. 4, Fig. S8, Tables S5, and S6, respectively).

The general quantitative distribution of the functional categories between Pool1 and Pool2 was similar. Notably, the most abundant KEGG category for both samples was carbohydrate metabolism (15% and 16% in Pool1 and Pool2 respectively) (Fig. 4 and Table S6). By contrast, the number of ORFs assigned to the COG class general function prediction in Pool2 was higher than in Pool1 (18% *vs* 15%, respectively), possibly reflecting the high number of unknown genes in NCBI for Pool1 (Table S5).

Interestingly, Pool2, if compared to Pool1, showed a significantly higher number of genes in the lipid metabolism and glycan biosynthesis KEGG categories (Fig. 4, Table S6). On the other hand, a much higher number of genes annotated in cell motility were detected in Pool1 (1.4%) than in Pool2 (0.5%) (Fig. 4, Table S6). In particular, the genes annotated in the flagellar assembly were remarkably more abundant in Pool1 (inset Fig. 4). The different nature of the two

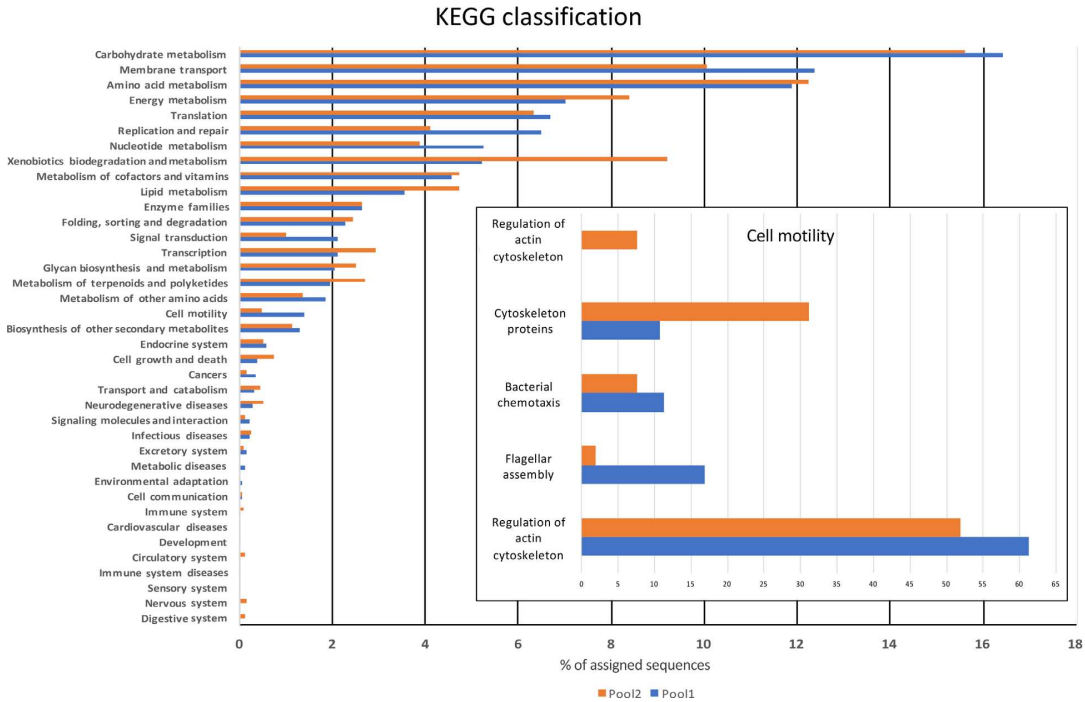


Fig. 4. KEGG analysis of the metagenome of solfatara Pisciarelli. Pool1 and 2 are compared according to KEGG functional categories. The inset shows the comparison of subcategories of cell motility.

samples might account for the bias in the functional annotation between them. Indeed, metagenome from Pool1 was isolated from a floating sample of muddy sediments, gravel, and water, while the sample from Pool2 was scraped from a side of the pool below the water level. Therefore, a large number of genes involved in lipid and glycan metabolism of the latter is consistent with a sessile lifestyle of the microbial population and consequently with active biofilm synthesis/degradation. In this regard, it has been demonstrated that the main genera populating Pool2, that is, *Metallosphaera, Sulfolobus,* and *Acidianus* are able to form active biofilms on different solid substrates, including metal sulfides [36] and synthetic Martian Regolith [37]. Instead, sampling in Pool1 might have picked up microorganisms whose metabolism relies on free motility in the aqueous environment.

The most striking differences between the two metagenomes were observed in the xenobiotics biodegradation and metabolism of terpenoids and polyketides categories (Fig. 4). Within these categories, which prevailed in Pool2, the two metagenomes showed xenobiotic pathways specific for each pool (Fig. S9). In particular, in Pool1 we found sequences putatively involved in the degradation pathways of trinitrotoluene and benzoate that are of potential interest for biotechnological processes such as bioremediation of (nitro)aromatic compounds [38–41]. In addition, regarding the benzoate degradation, both the pathways of CoA ligation (mainly represented by succinate dehydrogenases and acetyl-CoA C-acetyltransferases (EC 1.3.5.1–2.3.1.9) and hydroxylation (mainly acetyl-CoA acyltransferase (EC 2.3.1.16)) were identified, consistent with the presence in Pool1 of both aerobic and anaerobic microorganisms [42].

The functional annotation of the ORFs obtained by the assembly of the 'not assigned reads', by using KEGG and SEED [43] databases, still showed, as expected, the prevalence of the unclassified category. Carbohydrates and amino acid metabolism categories are the most abundant in both samples although, noticeably, specific differences in the distribution of the classes were found in the SEED classification. Indeed, Pool1 shows a significantly higher relative percentage of putative ORFs related to DNA metabolisms (7%), membrane transport (6%), nucleosides and nucleotides (6%), and sulfur metabolism (3%). By contrast in Pool2 prevalence of categories of carbohydrate (14% *vs* 12%, in Pool2 and Pool1, respectively), cofactors (12% *vs* 11%), respiration (11% *vs* 9%), amino acids (10% *vs* 9%), and protein metabolisms (7% *vs* 5%) was observed (Figs S10 and S11).

Carbohydrate-active enzymes

The highest number of sequences functionally annotated in the KEGG categories belonged to carbohydrate metabolism. Indeed, the solfatara Pisciarelli is surrounded by rich vegetation mainly composed of ferns, dicots, and grass, which contains starch as energy storage and, in their primary and secondary cell walls, specific hemicelluloses and pectins (Fig. 1C; Table S7). This observation prompted us to investigate on the different classes of CAZymes present in the two samples by comparing the amino acid sequences identified in Pool1 and Pool2 samples to the CAZy database (www.cazy.org) (Fig. 5; Tables 1, S8 and S9).

A total of 278 and 308 putative CAZymes were identified in Pool1 and 2, respectively, representing, in total, ~ 2.0% of all ORFs. It has been reported that CAZymes range between 1% and 5% of all the ORFs in all taxa genomes [20]. However, no survey of CAZymes from the microbial communities populating hydrothermal vents is currently available. Nevertheless, if compared to the much more complex human gut metagenome, in which the ORFs putatively encoding CAZymes corresponded to 2.6% of total ORFs, this value is remarkably high [44]. This suggests that the solfatara Pisciarelli represents a unique ecological niche for the discovery of new thermophilic CAZymes, for effective lignocellulose saccharification in biorefineries biotechnological applications.

Glycosyltransferases

Among CAZymes, the most abundant enzymatic class found in these metagenomes was glycosyltransferases (GTs), which represent about 60% and 56% of the putative CAZymes in Pool1 and 2, respectively (Fig. 5; Table 1). The most of GTs in both pools belonged to GT2 and GT4 families, which, in CAZy database include more than 200 000 entries, belonging to three domains of life, and group many activities (> 14 and > 20 different EC numbers in GT2 and GT4, respectively). Presumably, these two families account for the main carbohydrate-building activities of the microbial populations of the two Pools, including synthesis of α- and β-glycans, glycoconjugates, cell envelope glycans, exopolysaccharides, etc. [46]. In the two metagenomes, only a few other GT families have been identified. These include GT5 and GT35, whose archaeal members are glycogen synthases and phosphorylases, respectively [47], and GT66, which in Archaea are dolichyl-diphosphooligosaccharide protein glycotransferases involved in protein glycosylation. Remarkably, GT66 is essential in *S. acidocaldarius* [48]. The last

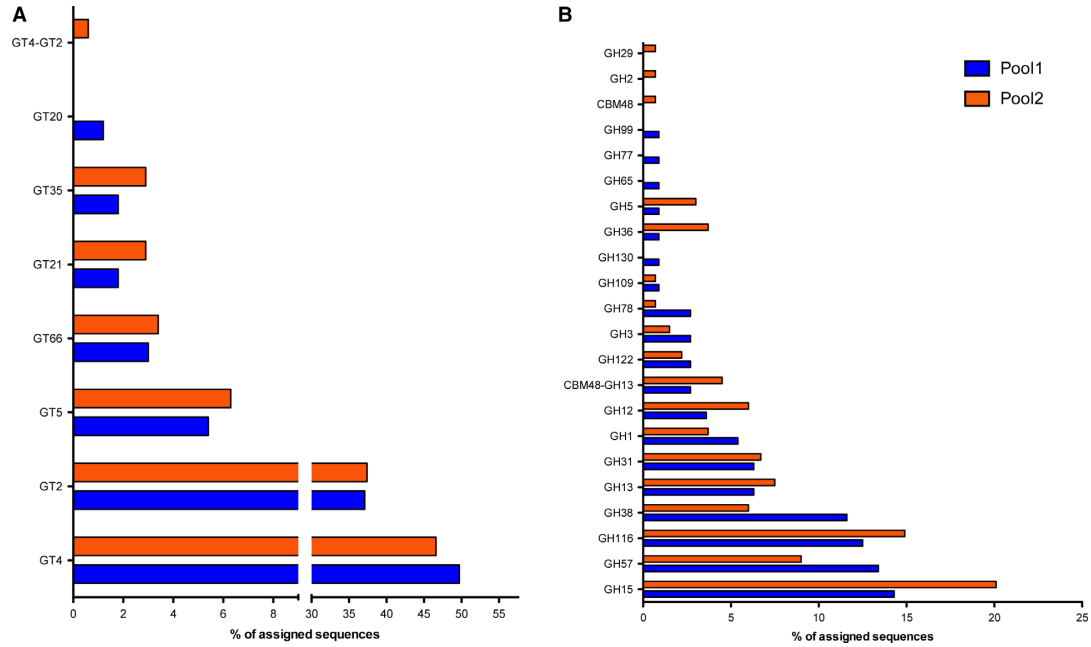


Fig. 5. Distribution of CAZymes. The percentage of assigned sequences of GTs (A) and GH and CBM (B) are displayed. CBM48 indicates sequences exclusively assigned to carbohydrate-binding module 48. CBM48-GH13 indicates sequences assigned to GH13 showing also a CBM48. CBM, CE, GH, GT.

two families are GT21, showing ceramide β-glucosyltransferases (EC 2.4.1.80) that have never been characterized in Archaea, and GT20, which could be identified only in Pool1 as homolog to a bifunctional trehalose-6-phosphate synthase/HAD hydrolase from *P. arsenaticum*. The study of GTs from hyperthermophiles is still in its infancy, essentially because these elusive enzymes are often membrane-associated, difficult to purify and characterize for their high specificity for donor and acceptor substrates. However, this metagenomic survey may offer access to diversified and novel hyperthermophilic GT biocatalysts.

Glycoside hydrolases and carbohydrate-binding modules

The second most abundant CAZymes identified, as 36% and 38% of Pool1 and 2, respectively, are putative GHs, followed by CE (2.5% and 3.2%), CBM (1.1% and 2.3%), and redox auxiliary activities (AA, 0.4% only in Pool1) (Tables S8 and S9). GH, AA, CE, and CBM families present in the two pools and the putative enzymatic activities associated so far according to CAZy database are summarized in Table 1. Among GHs, the GH1, GH2, GH3, GH5, GH10, GH12, GH29, GH78, GH116, GH130 families are hypothetically active on substrates containing equatorial β-D- and α-L-glycosidic linkages while GH4, GH13, GH15, GH31, GH36, GH38, GH57, GH65, GH77, GH99, GH109, GH122, are specific for axial α-D-glycosidic bonds. Interestingly, 14 enzymes found in Pool1 and 2 showed 100% identical amino acid sequence similarity to characterized glycosidases from *S. solfataricus* strains, namely, a β-glycosidase (EC 3.2.1.21) (GH1), a β-xylosidase/α-L-arabinosidase (EC 3.2.1.37/3.2.1.55) (GH3), a cellulase (EC 3.2.1.4) and a xylanase (EC 3.2.1.151) (GH12), a maltooligosyltrehalose synthase (EC 5.4.99.15) and a glycogen debranching enzyme (EC 2.4.1.18) (GH13), a glucoamylase (EC 3.2.1.3) (GH15), an α-L-fucosidase (EC 3.2.1.51) (GH29), an α-glucosidase (EC 3.2.1.20) and an α-xylosidase (EC 3.2.1.77) (GH31), an α-mannosidase (EC 3.2.1.24) (GH38), an α-amylase (EC 3.2.1.1) (GH57), and a β-glyco/xylosidase (EC 3.2.1.21/3.2.1.37) and a β-*N*-acetylglucosaminidase (EC 3.2.1.52) (GH116). These activities are involved in the hydrolysis and removal of sugar appendages of (hemi)-cellulose polysaccharides, in starch/glycogen mobilization, and in the turnover of the oligosaccharides of *N*-glycosylated proteins [49–65]. On the other hand, several GH families in Pool1 and 2 (GH2, GH65, GH78,

Table 1. CAZymes in solfatara Pisciarelli.

Family[a,b]	Pool1	Pool2
GT2	62	65
GT4	83	81
GT5	9	11
GT20	2	0
GT21	3	5
GT35	3	5
GT66	5	6
GH1	6	5
GH2	0	1
GH3	3	2
GH4	2	0
GH5	1	4
GH12	4	8
GH13	7	10
GH15	16	27
GH29	0	1
GH31	7	9
GH36	1	5
GH38	13	8
GH57	15	12
GH65	1	0
GH77	1	0
GH78	3	1
GH99	1	0
GH109	1	1
GH116	14	20
GH122	3	3
GH130	1	0

[a]From CAZy classification (www.cazy.org) on May 2019. Families with no previously characterized members in Archaea are highlighted in bold [20]. [b]For a detailed description of the activities in each family, please visit CAZypedia (www.cazypedia.org) [45].

GH99, GH130 in Table 1) have no known characterized archaeal homologs. In addition, we identified for the first time in Archaea a putative gene annotated as a member of GH109, a family previously only including α-*N*-acetylgalactosaminidase activities (EC 3.2.1.49). Instead, only archaeal members formed family GH122 where a single α-glucosidase has been characterized so far [66].

Putative CBM48 were identified in both samples associated with the C-terminal of GH13 α-amylases [56]. We found very few CBMs, which are remarkably underrepresented in Archaea (17 families out of > 80), no pectinases, and few GH families of *endo*-acting enzymes, namely GH5, GH12, and GH13. Possibly, in Pisciarelli Solfatara, the limited biodiversity of the plants, containing only certain (hemi)celluloses (Table S7), did not favor the enrichment of the microbial community with a variety of polysaccharide-degrading strains. Moreover, the very hot and acidic environment is able to efficiently extract pectins and increase the accessibility of the plant polysaccharides coming from the biomass falling into the pools. It is worth noting, indeed, that while low pHs in the range of 1.0–3.0 with high acid concentrations can allow the hydrolysis of different polysaccharides such as pectins (pH 1.5 at 70 °C), softwood hemicelluloses (pH 1.5 at 150 °C in 60 min), arabinogalactans (pH 1.0 at 90 °C in 24 h), and galactoglucomannan (in the pH 1.0–3.0, at 50–90 °C), the oligomers formation is more prominent in the presence of dilute acids with higher pHs value [67–71]. The released oligosaccharides, once internalized into the cells, can become substrates for the GHs. Therefore, the chemico-physical conditions of the hydrothermal vent produced did not exert environmental pressure to maintain endoglycanases, CBMs, and AAs in this microbial community (see also below). On the other hand, the CAZymes found in this microbial community might be of special interest because of their ability to hydrolyze recalcitrant polysaccharides and withstand both temperature and pH extremes.

Certain CAZyme families were specific for Pool1, namely GH4, GH65, GH77, GH99, and GH130, while enzymes from families GH2, GH29, and CBM48, which were only found in Pool2, showed the highest identity to *S. solfataricus* strain P2, which is also present in Pool1 (Fig. 5B; Table S8). Presumably, we could not find these families in Pool1 because the encoding regions were not covered by sequencing. Instead, members of GH families characteristic of Pool1 showed high similarities to those from organisms which are absent in Pool2, confirming their uniqueness. The harsh conditions induced a specific selection of the microorganisms populating Pisciarelli reducing the biodiversity of Pool2. However, the high number of sequences not matching any known sequence in NT database may account for unknown species populating this site.

While these hypotheses are currently under study, to shed light on the *CAZome* of Pool2, we produced and characterized two enzymes belonging to two different contigs of cluster 9, which, from the *in silico* analysis may originate from a novel microorganism not yet identified and belonging to the family of *Sulfolobaceae* (see above). In particular, we selected putative enzymes from the GH5 and GH109 families. The former includes many glucanases involved in polysaccharide hydrolysis and could be of great interest for biotechnological applications such as in the second-generation biorefineries, while in GH109, which are particularly interesting for the removal of α-GlcNAc from the antigens from red blood cells [72,73], members from Archaea have never been identified.

GH5_Pool2

Cloning, expression, and purification

The GH5 family is one of the largest in the CAZy database, with 13 792 entries in May 2019, and members widely distributed across Archaea, Bacteria, Eukarya, and viruses. A variety of substrate specificities are associated with enzymes belonging to this family, notably endoglucanase (EC 3.2.1.4) (cellulase) and endomannanase (EC 3.2.1.78), as well as exoglucanases (EC 3.2.1.91), exomannanases (EC 3.2.1.100), β-glucosidase (EC 3.2.1.21), and β-mannosidase (EC 3.2.1.25). A further classification allowed the division of GH5 into 56 distinct subfamilies [74]. In May 2019, there are 563 characterized GH5 and only 5 of them are from Archaea: two cellulases, from *Desulfurococcaceae archaeon EBI-244* and *Halorhabdus utahensis* DSM 12940, two endo-β-1,4-glucanases, from *Pyrococcus horikoshii OT3* and *P. abyssi GE5*, and the endoglucanase Cel5A from an uncultured archaeon [75–79]. Family GH5 enzymes are retaining enzymes that follow a classical Koshland double-displacement mechanism and present a glutamic acid and a glutamate as catalytic residues [80,81]. Structurally, GH5 enzymes show a classical $(\alpha/\beta)_8$ TIM barrel fold.

The GH5 enzymes from Archaea mainly belong to subfamilies 1 and 19 (GH5_1 and GH5_19, respectively) in the CAZy classification; GH5_Pool2 is a member of subfamily GH5_19 in which only the bacterial β-mannosidase from *Thermotoga thermarum* was characterized [82]. Its gene was identified in the contig_5232 of Pool2 belonging to the cluster 9. The region surrounding GH5_Pool2 sequence (Fig. 6A) encodes for a hypothetical Major facilitator superfamily (MFS) transporter, a hypothetical alcohol dehydrogenase, and an oxidoreductase with good identity to the sequences of *Candidatus acidianus copahuensis* (74%), *S. tokodaii* (75%), and *Sulfolobus islandicus* (67%), respectively. However, this contig encodes for 179 hypothetical proteins (> 200 aa) with identity between 28% and 89% toward sequences belonging mainly to the order of Sulfolobales (Table S10), suggesting that it could belong to the chromosome of a new, unclassified archaeon.

The GH5_Pool2 gene encodes for a protein of 620 amino acids with a hypothetical molecular mass of about 71.3 kDa. The analysis of GH5_Pool2 amino acid sequence by ClustalO multialignment with the characterized GH5 β-mannosidase from *T. thermarum* and its homologs showed the conservation of the catalytic residues, suggesting that the GH5_Pool2 enzyme could be functional (Fig. S12).

The analysis by BlastP of the predicted gene product of GH5_Pool2 showed low homology with characterized enzymes. The highest amino acid sequence identity was as low as 58% with a hypothetical protein from *Sulfolobus* sp. A20 (WP_069282934.1) and revealed a conserved domain belonging to the endo-1,4-β-mannosidases (COG3934) between residue Thr2 and Gly590, while the analysis of the sequence using the hidden Markov model indicates a match with GH5_19 from the Ile5 residue to Lys274. To date, no such activity has been reported among the characterized archaeal GH5 enzymes.

The coding sequence of GH5_Pool2 was amplified by PCR from Pool2 metagenomic DNA and cloned and expressed in *Escherichia coli* as described in the Materials and methods. The recombinant protein was purified through two simple steps with a final yield of 65.3% (2.2 $mg \cdot L^{-1}$ of culture) and a purity of 95%.

Molecular mass determination and effect of temperature and pH

The recombinant GH5_Pool2, analyzed by SDS/PAGE, revealed the presence of a single band of the expected molecular weight of 73.4 kDa, corresponding to a monomer (Fig. S13, Table S11). Other thermophilic members of GH5 are reported to be oligomers in native form. Indeed, the closest characterized

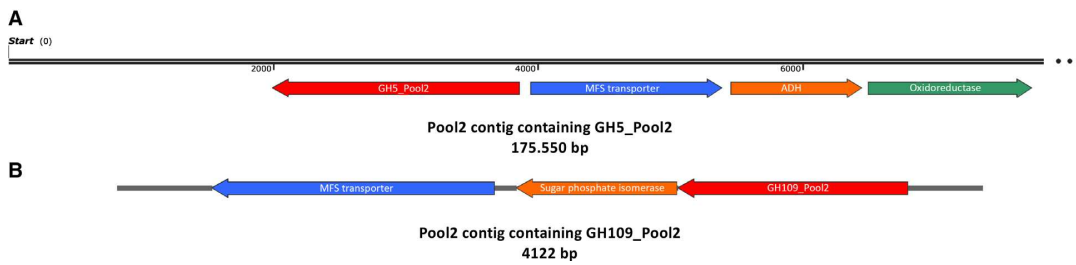


Fig. 6. The genomic environment of GH5_Pool2 and GH109_Pool2. (A) Close-up of the contig_5232 containing the gene encoding GH5_Pool2. (B) Contig_2182 containing the gene of GH109_Pool2.

homolog from *T. thermarum* β-mannosidase (TthMan5) is a heptamer in solution [82]. The molecular mass of the native GH5_Pool2 was 271 ± 13.5 kDa, as determined by size-exclusion chromatography (Fig. S14), indicating that the enzyme is a tetramer in solution.

GH5_Pool2 is optimally active on 4-Nitrophenyl β-D-mannopyranoside (4NP-β-Man) at pH 5.5 while maximum activity (~ 0.40 U·mg^{-1}) was maintained at temperatures from 65 to 85 °C. (Fig. 7), indicating a temperature optimum broader than that of the TthMan5 β-mannosidase, showing a peak at 85 °C [82]. These data confirm the thermophilic nature of GH5_Pool2, as expected from the harsh conditions of Pool2. However, the enzyme was only barely active at the pH < 5 recorded in Pool2, thereby excluding the enzyme is secreted, as also supported by SignalP analysis indicating the absence of a signal peptide (data not shown).

(Hyper)thermophilic enzymes are exceptionally temperature stable. Notably, the enzyme was 100% active on 4NP-β-Man for more than 22 h at 65 and 85 °C and, at 95 °C, it maintained 50% of activity for 4 h (Fig. 7C). These results confirmed that GH5_Pool2 is an extremely thermostable enzyme. It is worth noting that such high-temperature stability was not reported for the other five archaeal GH5. Indeed, TthMan5 β-mannosidase is 91% and 50% active after 2 h at 75 and 90 °C, respectively [82]. Other β-mannosidases from hyperthermophiles share similar lower thermostability, such as those from *T. neapolitana* (2 h 90 °C) and *T. maritima* (4 h at 80 °C) [76,83]. Again, the high thermostability of GH5_Pool2 is consistent with the harsh conditions found in Pool2.

Substrate specificity

To investigate the substrate specificity of GH5_Pool2, kinetic parameters for hydrolysis on different substrates were determined at 65 °C (Table 2), revealing a k_{cat} of 32.9 s^{-1}, 5.70 s^{-1}, and 0.76 s^{-1} toward mannopentaose, 4NP-β-Man, and 4NP-β-Glc, respectively. According to these results, GH5_Pool2 should be classified as a β-mannanase. Moreover, GH5_Pool2 showed substrate inhibition with 4NP-β-Man concentrations higher than 0.175 mM (Fig. S15A). This is not uncommon in glycosidases, which frequently have long active site clefts to accommodate multiple monosaccharide units of a polysaccharide chain and, thus, have the potential to bind two substrate molecules simultaneously in an unproductive fashion [84]. For this reason, the kinetic constants were determined by linear regression of the Lineweaver–Burk plot for substrate concentrations between 0.025 and 0.15 mM (Fig. S15B).

GH5_Pool2 was assayed on different aryl- glycosides, oligo-, and polysaccharides to determine the substrate specificity in more detail. The enzyme showed the highest specific activity on mannopentaose, followed by mannotetraose and laminaritetraose. Low activity was found on 4NP-β-Man, 4NP-β-Glc, 4-Nitrophenyl β-D-galactopyranoside (4NP-β-Gal), and laminarin, whereas no activity was detected either on the other aryl-glycosides tested, on cellulose oligosaccharides or guar galactomannan (data not shown). Additionally, GH5_Pool2 catalyzed the partial hydrolysis of 6^3,6^4-α-galactosyl-mannopentaose (Fig. S16, lanes 1 and 2) and the complete hydrolysis of

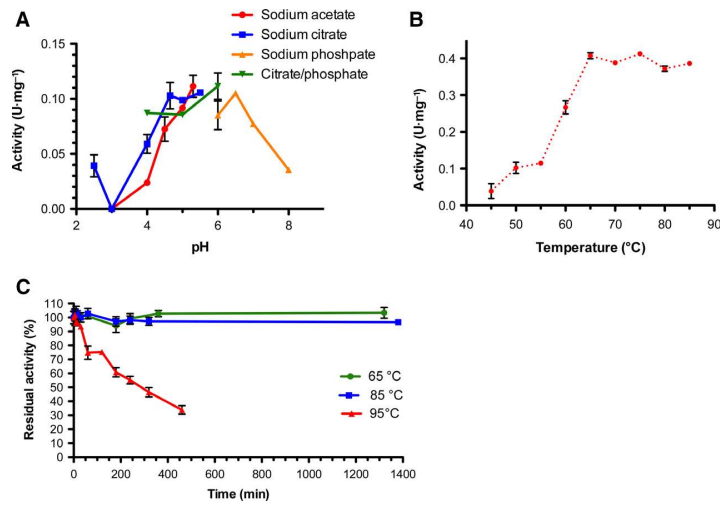


Fig. 7. Biochemical characterization of GH5_Pool2. (A) pH dependence, (B) temperature dependence, and (C) thermostability at 65, 85, and 95 °C.

Table 2. Steady-state kinetic constants of GH5_Pool2 on different substrates.

Substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($mM^{-1}·s^{-1}$)
Mannopentaose	3.13 ± 1.20	32.90 ± 6.03	10.5
4NP-β- Glc	9.90 ± 2.40	0.76 ± 0.06	0.076
4NP-β-Man	0.42 ± 0.07	5.70 ± 0.45	13.75

mannotetraose (Fig. S16, lane 3 and 4) and mannopentaose (Fig. S16, lanes 5 and 6), as previously shown for other GH5 β-1,4-mannanases [85].

As reported above, the enzyme showed activity on different laminarin oligosaccharides. For this reason, the ability of GH5_Pool2 to hydrolyze laminaribiose (Glc2), laminaritriose (Glc3), laminaritetraose (Glc4), laminaripentaose (Glc5) (Fig. S17) was further investigated by HPAEC-PAD (Figs S18-S21) revealing that the enzyme is *exo*-acting. In fact, in all cases, the chromatograms showed as products glucose and a polysaccharide shortened by one unit if compared to the substrate. These results clearly indicate that GH5_Pool2 shows β-mannanase and glucan β-1,3-glucosidase activity. This is not surprising as all but one member of characterized enzymes belonging to GH5_19 subfamily are *exo*-acting on laminarin. GH5_Pool2 is the first thermophilic member from this subfamily, possibly belonging to the Archaea domain and showing β-mannanase activity. The plants present in solfatara Pisciarelli contain gluco- and galactomannans (Table S1), suggesting that GH5_Pool2 could be involved in the degradation of oligomers from this source. Regarding the β-1,3-glucanase activity, the enzyme might be involved in the degradation of the hemicellulosic β-1,3-β-1,4-glucan and callose from the plants, or in the turnover of glycans produced by the extremophilic microorganisms living Pool2. The observed optimal activity at pH close to neutrality, the absence of a signal peptide for secretion, and the presence of a hypothetical MFS transporter gene flanking the gene encoding for GH5_Pool2, suggest that the enzyme is active on oligosaccharides once internalized into the cytoplasm.

GH109_Pool2

Cloning, expression, and purification

As of May 2019, the CAZy database lists 713 family GH109 entries distributed exclusively among Bacteria with α-*N*-acetylgalactosaminidase as the only activity reported. To date, only 6 members of the family have been characterized, three of them from *Elizabethkingia meningoseptica* strains. Other members of the GH109 family did not exhibit α-*N*-acetylgalactosaminidase activity, suggesting that the family may include enzymes with different specificities [72].

GH109 enzymes do not follow the classical Koshland double-displacement mechanism [86], but instead, use a rare catalytic mechanism involving NAD-dependent hydrolysis similar to the one reported in the GH4 family [87]. The enzymes from the latter family are unique in their ability to cleave both α- and β-linked glycosides. The reaction mechanism of GH4 enzymes involves the reduction of NAD^+ cofactor to NADH and oxidation of the substrate at the C3 hydroxyl group with the formation of an unsaturated intermediate. A catalytic base deprotonates the C2 and the O-glycosidic linkage is cleaved by α, β-elimination. The addition of a water molecule at C1 and the reduction of the C3 by the same NADH molecule generate the hydrolysis product and the initial NAD^+. The GH4 β-glucosidase from *T. maritima* is activated by 1 μM NAD^+, 10 mM mercaptoethanol, and 1 mM Mn^{2+} [87]. GH109 enzymes most likely operate with the same mechanism; however, in enzymes characterized so far, NAD^+ is embedded in a narrow active site tunnel devoided of metal ions and they did not display either metal requirements or higher activity by addition of exogenous NAD^+ [72].

The GH109_Pool2 gene was identified at the far end of the contig_2182 of Pool2, belonging to cluster 9 as well as the one containing GH5_Pool2 (contig_5232). The whole contig encodes only three hypothetical proteins (Fig. 6B) including a hypothetical sugar phosphate isomerase and an MFS transporter homolog to the sequences from *Sulfolobus* sp. A20 (83%) and *Candidatus bathyarchaeota archaeon* (47%), respectively. The GH109_Pool2 gene encodes for a protein of 364 amino acids with a calculated molecular mass of about 41.26 kDa, which makes it smaller than the GH109 enzymes characterized to date (about 50 kDa). However, GH109_Pool2 showed the conserved domain of the GH109 family at the N-terminal region (from Ile2 to Arg115). The deduced protein sequence was analyzed through ClustalO multialignment with other members of the family revealing that it was distantly clustered from a characterized α-*N*-acetylgalactosaminidase (EC 3.2.1.49) (Fig. S22).

GH109_Pool2 was produced in *E. coli* as a fusion protein with a 6x(His) tag, but, since it led to an insoluble protein (data not shown), it was subsequently produced as a fusion to an *N*-terminal tag of NusA [88] followed by a 6x(His) tag. The recombinant protein was purified by IMAC and a size-exclusion chromatography yielding 0.3 $mg·L^{-1}$ of 95% pure protein

(Fig. S23, Table S12). The single band observed by SDS/PAGE showed a molecular mass of about 98.43 kDa corresponding to the monomer of GH109_Pool2 linked to NusA; the molecular mass of 776.24 ± 1.7 kDa in native conditions indicated that the enzyme was an octamer in solution. (Fig. S24).

Substrate specificity

Reaction mixtures of GH109_Pool2, containing different 4NP-glycosides (5 mM) and incubated at 75 °C in 50 mM sodium phosphate buffer pH 8.0, revealed that the enzyme was active on aryl-glycosides in both α- and β-configuration. In particular, the enzyme was able to hydrolyze 4-Nitrophenyl α-D-glucopyranoside (4NP-α-Glc) (0.33 ± 0.01 U·mg^{-1}), 4NP-β-Glc (0.59 ± 0.01), 4-Nitrophenyl *N*-acetyl-α-D-glucosaminide (4NP-α-GlcNAc) (0.52 ± 0.01 U·mg^{-1}), and 4NP-β-GlcNAc (0.75 ± 0.02 U·mg^{-1}) while it was more than tenfold less active on 4NP-β-Man (0.019 ± 0.001 U·mg^{-1}) and 4NP-β-Gal (0.018 ± 0.001 U·mg^{-1}). By contrast, no activity was observed on 4-Nitrophenyl *N*-acetyl-β-D-galactosaminide, 4-Nitrophenyl α-D-galactopyranoside (4NP-α-Gal), 4-Nitrophenyl α-D-mannopyranoside, and especially 4NP-α-GalNAc, which is the only known substrate of GH109 enzymes. The enzyme was then assayed at 85 °C in optimal conditions in the presence of divalent metals, metal-chelating agents, and DTT (50 mM) (Fig. 8A). EDTA and DTT, as well as $FeCl_3$ and $MnCl_2$, did not significantly influence the activity. On the contrary, 1 mM $CuCl_2$ completely inhibited the enzyme similar to previous results on the α-*N*-acetylgalactosaminidase from *Elizabethkingia meningosepticum* [72].

The hydrolytic activity displayed by GH109_Pool2 occurred exclusively in the presence of 1.5 mM NAD^+ similarly to the other GH109 enzymes characterized so far [72] To investigate the dependence of NAD^+, the enzyme was assayed on 4NP-β-Glc (5 mM) at different NAD^+ concentrations showing an activation at 0.05 mM with a maximum of activity at 0.5 mM and a K_d of 32 ± 5 μM confirming that NAD^+ is required in the GH109 family (Fig. 8B).

To investigate the substrate specificity of GH109_Pool2, the steady-state kinetic constants of 4NP-α and β-glucosides and *N*-acetyl-glucosides were determined at 75 °C (Table 3). The k_{cat} values were similar among all the tested substrates, but the lower K_M values measured on 4NP-β-anomers (0.12 mM) made of 4NP-β-GlcNAc, 4NP-β-Glc the preferred substrates.

GH109_Pool2 was optimally active in sodium phosphate buffer at pH 8.0 and at 85 °C on 4NP-β-Glc (1.44 ± 0.04 U·mg^{-1}), confirming that it was thermophilic (Fig. 9A,B). In addition, the enzyme maintained up to 40% activity after 1 h at 75 °C while it was quickly inactivated when incubated at 85 °C (25% residual activity after 5 min) (Fig. 9C).

The kinetic characterization showed that this firstly identified archaeal enzyme of the GH109 family differs significantly from the other known members. Indeed, GH109_Pool2 is the first characterized β-*N*-acetylglucosaminidase/β-glucosidase (EC 3.2.1.52/3.2.1.21), expanding the activity repertoire of this CAZy family, it is activated by exogenous addition of NAD^+, and cleaves both α- and β-linked substrates, two characteristics only known so far in GH4 enzymes [89], suggesting structures and functional analogies never observed before between the GH4 and GH109 families.

Table 3. Steady-state kinetic constants of GH109_Pool2 on different substrates.

Substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM $^{-1}$·s^{-1})
4NP-β-Glc	0.12 ± 0.01	4.75 ± 0.004	39.58
4NP-β-GlcNAc	0.12 ± 0.01	6.11 ± 0.01	50.91
4NP-α-Glc	1.48 ± 0.18	3.28 ± 0.02	2.22
4NP-α-GlcNAc	0.87 ± 0.06	4.98 ± 0.01	5.72

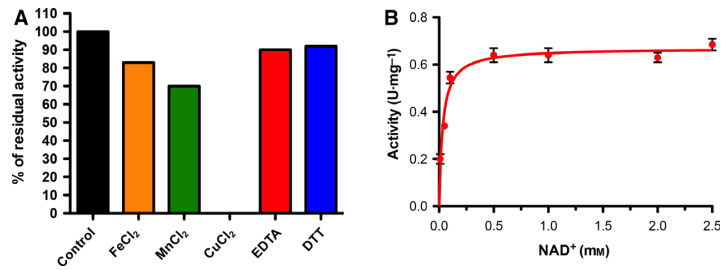


Fig. 8. Effect of additives and cofactors on GH109_Pool2. (A) Effect of different additives on enzyme activity: (black bar) control, (orange bar) $FeCl_2$ 1 mM, (green bar) $MnCl_2$ 1 mM, (red bar) EDTA 1 mM, and (blue bar) DTT 50 mM. (B) NAD^+ dependence of GH109_Pool2.

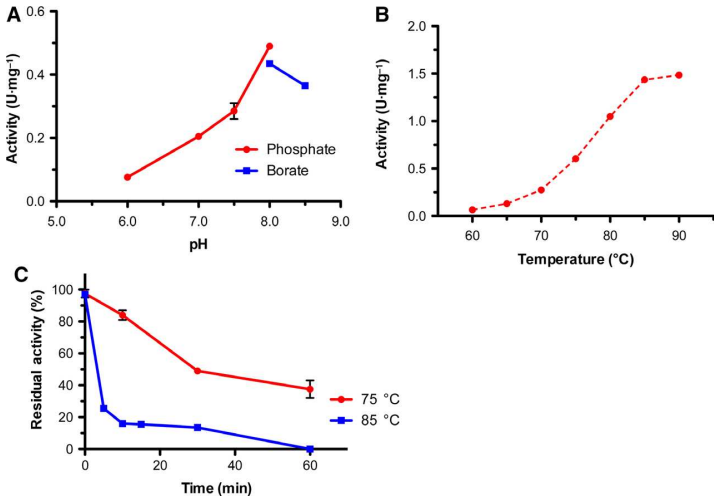


Fig. 9. Biochemical characterization of GH109_Pool2. (A) pH dependence in sodium phosphate (red) and sodium borate (blue) buffers. (B) Temperature dependence. (C) Thermostability at 75 °C (red) and 85 °C (blue).

Conclusions

The metagenomic investigation of the solfatara Pisciarelli revealed a rich *CAZome*, offering the opportunity to discover CAZy with new substrate specificities, reaction mechanisms, and optimal catalytic conditions. Only a minority of these putative CAZymes (about 14% and 22%, in Pool1 and 2, respectively) is 100% identical to those available in the *NR* database, which, for the most part, have never been characterized and mostly result from annotated (meta)genomes. In addition, 24% and 20% of CAZymes in Pool1 and 2, respectively, show 25–60% identities to sequences in the data banks. Considering that even single amino acid substitutions can substantially modify enzymatic activity, substrate specificity, stability, and other properties, virtually all CAZymes identified in Pool1 and 2 require a detailed enzymatic characterization to confirm the inferred activity. The expected intrinsic stability to high temperatures and very acidic pHs of Pool1 and 2 enzymes makes this survey a valuable tool for future exploitation in biotechnology. In particular, second-generation biorefineries, which often require very hot and acidic conditions, are an ideal field of application not only of CAZymes but also of other industrial enzymes.

This study demonstrates that even sites that have been consistently sampled for decades are still largely unexplored in terms of microbial diversity. To the best of our knowledge, this is the first *CAZome* study of a hydrothermal vent in which CAZymes have been systematically annotated, their role predicted on the basis of the source of lignocellulose in the natural environment, produced, and characterized in detail. This approach revealed CAZyme families previously unknown in Archaea, putative enzymes with promising potential for utilization in next-generation biorefineries, and novel GHs with interesting substrate specificities and stability.

Materials and methods

Solfatara sampling

Samples from the hydrothermal mud/water pools, Pool1 and Pool2, respectively, were transferred into sterile serum bottles, respectively, were closed and immediately transferred to the laboratory for DNA extraction. *In situ* measurements of temperature and pH were performed by using a HI-93510 thermometer (HANNA instruments®, Padova, Italy) equipped with a Pt100 probe and litmus tests. Then, pH was accurately measured again with a pH meter (Crison Instruments, Inc., Barcelona, Spain) in the laboratory.

Isolation of DNA

DNA of the original samples was extracted with Power Soil DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA) by following the manufacturer's protocol.

Metagenomic sequencing and *in silico* analysis

The extracted and purified metagenomic DNA of each of the original samples was used for shotgun sequencing with HISeq2000 (Illumina) performed at Beijing Genomics Institute (BGI-Shenzhen) Shenzhen, China. Raw sequencing datasets were processed at the Institute of Biosciences and BioResources, CNR, Naples, Italy. Adapter removal and

trimming of bad quality sequencing parts within the raw metagenomic reads (33 333 332 for Pool1 and 26 666 666 for Pool2) were performed by using Trimmomatic [90] prior to *de novo* assembly.

Clean reads (33 240 694 and 26 507 584 for Pool1 and Pool2, respectively) were assembled using MEGAHIT [91] by using kmer of 33 for Pool1 and 47 for Pool2. The assembly of Pool1 was composed by 2148 contigs (≥ 500 bp) with a total length of 7350828 bp, N50 = 27918 bp, N90 = 833 bp, max contig length = 365616 bp. The assembly of Pool2 was composed by 4263 contigs (≥ 500 bp) with a total length of 15668449 bp, N50 = 16634 bp, N90 = 1097 bp, max contig length = 241427 bp. Assembled contigs (≥ 1000 bp) were grouped into bins by using MyCC [34] to reconstruct the metabolic potential of representative genomes and the quality of the bins was analyzed by CheckM v1.0.12 [35].

For the functional annotation, all the contigs obtained with length ≥ 500 bp were analyzed by using Prodigal [92] and were clustered to remove the duplicate (100% identity) by using CD-HIT [93]. ORFs with the length ≥ 100 bp were analyzed by using Blast+/BlastP [94] against KEGG and COG databases with *E*-value < 1E-5.

For microbial diversity analysis including the calculation of the Shannon–Weaver index and the Simpson's reciprocal index, short paired-end Illumina reads (90 bp) were aligned to the nucleotide reference database of NCBI *NT* by using Blast+/BlastN and the resulting output data of each sample (Pool1: reads = 22 844 457 with match in *NT*, assigned reads = 22 838 568; Pool2: reads: 10 445 671 with match in *NT*, assigned reads = 10 443 205) was used as input for MEGAN6 Community Edition with the following parameters: MinScore = 50.0, MaxExpected = 0.01, TopPercent = 10.0, MinSupportPercent = 0.05, mode = BlastN) [95].

The microbial diversity analysis based on the alignment of the short paired-end Illumina reads against the protein reference database of NCBI *NR* was performed by using DIAMOND [96]. Again, the resulting output data of each sample (Pool1: reads = 24 846 319 with match in *NR*, assigned reads = 24 842 956; Pool2: reads: 13 442 717 with match in *NR*, assigned reads = 13 428 467) were used as input for MEGAN6 Community Edition with the following parameters: MinScore = 35.0, MaxExpected = 0.01, TopPercent = 10.0, MinSupportPercent = 0.05, mode = BlastP).

Rarefaction curves for both samples were calculated by evaluating the k-mer redundancy of the reads with Nonpareil 3 [25].

For the identification of the enzymes active on carbohydrates, the ORFs identified in Pool1 and Pool2 were analyzed against CAZy database using BLASTp and hidden Markov models as described [20].

The reads showing no match against 'NT' database (10 638 917 and 16 410 389 for Pool1 and Pool2, respectively) were reassembled by using SOAPdenovo (kmer values between 23 and 63). The contigs obtained were clustered to remove duplicate (100% identity) by using CD-HIT [93] and aligned again to the *NT* database and analyzed by using Prodigal [92] for the identification of the coding sequences.

All the ORFs identified were analyzed by using Blast+/BlastP against the *NR* database, and the resulting output data of each sample were used as input for MEGAN6 Community Edition with the following parameters: MinScore = 35.0, MaxExpected = 0.01, TopPercent = 10.0, MinSupportpercent = 0.05, mode = BlastP) for the functional and taxonomical annotation.

Cloning of GHs from Pool2 metagenomic DNA

The genes encoding GH5_Pool2 and GH109_Pool2 were amplified by PCR from Pool2 metagenomic DNA and cloned into a pET-28a and pHTP10 vector, respectively, according to NZYTech protocols. The resulting vectors pET-28a-GH5_Pool2, containing a 6x(His) tag at the N-terminal region, and pHTP10-GH109_Pool2, allowing the expression of GH109_Pool2 in fusion with 6x(His) tag and NusA protein to promote solubility and folding, were used for the subsequent expression of the enzymes.

GHs expression and purification

Escherichia coli BL21(DE3)-RIL/pET-28a-GH5_Pool2 cells were grown in 2 L of LB at 37 °C supplemented with kanamycin (50 $\mu g \cdot mL^{-1}$) and chloramphenicol (30 $\mu g \cdot mL^{-1}$).

Gene expression was induced by the addition of 1.0 mM IPTG when the culture reached an A_{600} of 0.5. Growth was allowed to proceed for 16h, and cells were harvested by centrifugation at 6000 ***g*** for 30 min.

The resulting cell pellet was resuspended in 50 mM sodium phosphate buffer, pH 7.3, 300 mM NaCl, and Triton 1% (v/v) in a ratio of 1 : 5 (w/v), incubated with DNAseI (1 $mg \cdot mL^{-1}$) (Sigma-Aldrich, St. Louis, MO, USA) and lysozyme (1 $mg \cdot mL^{-1}$) (Sigma-Aldrich) for 1 h at room temperature, and homogenized by French cell pressure treatment. After centrifugation for 30 min at 11 000 ***g*** (4 °C), the free cell extract was loaded on Protino® Ni-TED Packed Columns (Macherey-Nagel, Düren, Germany) according to the manufacturer. The eluted fractions were pooled, dialyzed against PBS 1X, pH 7.3, and heat-fractioned at 80 °C for 30 min. Finally, to remove thermally denaturated proteins, the sample was centrifuged at 16 000 ***g*** for 30 min at 4 °C.

Escherichia coli One Shot™ BL21 Star™ (DE3)/pHTP10-GH109_Pool2 cells were grown in 2 L of LB at 37 °C supplemented with kanamycin (50 $\mu g \cdot mL^{-1}$). Gene expression was induced by the addition of 1 mM IPTG when the culture reached an A_{600} of 0.5 at 37 °C. Growth was allowed to proceed for 16 h at 20 °C, and cells were harvested by centrifugation at 5500 ***g*** for 20 min. The resulting cell pellet was resuspended in 25 mM sodium phosphate buffer, pH

7.3, 150 mM NaCl, and Triton 1% (v/v) in a ratio of 1 : 5 (w/v), incubated with DNAseI (1 mg·mL^{-1}) (Sigma-Aldrich) and lysozyme (1 mg·mL^{-1}) (Sigma-Aldrich) for 1 h at room temperature. Cells were lysed by five sonication cycles, and cell debris was removed by centrifugation at 16 000 ***g*** for 30 min. The free cell extract was loaded on a HisTrap HP 1 mL (GE Healthcare, Chicago, IL, USA) in 25 mM sodium phosphate buffer, pH 7.3, 150 mM NaCl and eluted with imidazole 250 mM. The eluted fractions were pooled, dialyzed against PBS 1X pH 7.3, and loaded on a HiLoad Superdex 200 10/60 (GE Healthcare) by using an isocratic elution (1.0 mL·min^{-1} of PBS 1X pH 7.3).

GH5_Pool2 activity assay

The standard assay of GH5_Pool2 (30 μg, 0.55 μM) was performed in 50 mM sodium citrate buffer, pH 5.5 on 4NP-β-Man at the final concentration of 0.15 mM at 65 °C, in a final volume of 200 μL. After 2 min of incubation, the reaction was blocked in ice and by adding 800 μL of 1 M sodium carbonate pH 10.2. The absorbance was spectrophotometrically measured at 420 nm at room temperature; the mM extinction coefficient of 4-nitrophenol in this condition is 17.2 mM^{-1}·cm^{-1}. Spontaneous hydrolysis of the substrate was subtracted by using appropriate blank mixtures without the enzyme.

One enzymatic unit is defined as the amount of enzyme catalyzing the conversion of 1 μmol of substrate into product in 1 min, at the indicated conditions.

Enzymatic assays of GH5_Pool2 on oligosaccharides and polysaccharides

In order to determine the μmol of reducing ends produced by the enzyme, calibration curves were created by using known amount (i.e., 0.02, 0.04, 0.05, 0.10, 0.14 μmol) of either glucose or maltose. Then, the enzymatic activities on oligosaccharides (6^3, 6^4-α-galactosyl-mannopentaose, mannotetraose, mannopentaose, laminaribiose, laminaritriose, laminaritetraose, laminaripentaose) (1 mg·mL^{-1}), and on polysaccharides (laminarin and guar galactomannan) (1 mg·mL^{-1}) were analyzed by using the Somogyi–Nelson assay for reducing sugars as previously reported [97].

Effect of pH and temperature

The temperature dependence of the activity of GH5_Pool2 was determined in the range 45–85 °C in 50 mM sodium citrate buffer (pH 5.5) with 4NP-β-Man (0.15 mM) as substrate.

Optimal pH and buffered saline conditions were determined by assaying GH5_Pool2 with 20 mM 4NP-β-Glc at 65 °C for 2 min in 50 mM of sodium acetate buffer (pHs 3.0–5.5), 50 mM of citrate phosphate buffer (pHs 4.0–6.0), 50 mM citrate buffer (pHs 2.5–5.5), and in 50 mM phosphate buffer (pHs 6.0–8.0).

The temperature dependence of the activity of GH109_Pool2 (3 μg, 3.86 nM) was determined on 4NP-β-Glc (5.0 mM) with NAD^+ (1.5 mM) in the range 60–90 °C in 50 mM sodium phosphate buffer (pH 8.0) in a final volume of 1 mL.

Optimal pH of GH109_Pool2 was determined by assaying the enzyme (3 μg, 3.86 nM) with 5 mM 4NP-β-Glc, NAD^+ 1.5 mM at 75 °C in 50 mM of phosphate buffer (pHs 6.0–8.0), or 50 mM sodium borate buffer (pH 8.0–8.5) in a final volume of 1 mL. The absorbance was spectrophotometrically measured at 420 nm.

Thermal stability of GH5_Pool2

The thermostability of GH5_Pool2 was evaluated by incubating the enzyme in PBS 1X (pH 7.3) at 65, 85, and 95 °C up to 23 h. Aliquots (30 μg) were withdrawn at the indicated times and assayed at standard conditions at the final concentration of 0.55 μM. The residual activity was expressed as a percentage of the maximal enzymatic activity measured before the incubation at indicated temperatures.

Determination of kinetic parameters of GH5_Pool2

Kinetic constants of GH5_Pool2 were measured at standard conditions by using concentrations of substrate ranging between 0.01–0.5 mM for 4NP-β-Man, 0.5–40 mM for 4NP-β-Glc.

For mannopentaose, the range used is 0.25–8 mM and the enzymatic activity was detected by using the Somogyi–Nelson assay. All kinetic data were calculated as the average of at least two experiments and were plotted with the program GRAPHPAD PRISM (San Diego, CA, USA).

Specific activity assay of GH109_Pool2

The specific activity for GH109_Pool2 (3 μg, 3.86 nM) was measured in 50 mM sodium phosphate buffer, pH 8.0 at 75 °C, in the presence of 1.5 mM NAD^+, on 4NP-substrates at the final concentration of 5 mM in the final volume of 1 mL. The mM extinction coefficient for 4-nitrophenol at 420 nm at standard conditions and 75 °C was 19.1 mM^{-1}·cm^{-1}. One unit of enzyme activity was defined as the amount of enzyme catalyzing the hydrolysis of 1 μmol of substrate in 1 min at the conditions described.

Determination of the GH109_Pool2 dependence for NAD^+ as cofactor

The activity of GH109_Pool2 in presence of different NAD^+ concentration (0.01–5.0 mM) was measured by

assaying 3 μg of enzyme, at the final concentration of 3.86 nM, on 4NP-β-Glc 5, in 50 mM sodium phosphate buffer pH 8.0 at 75 °C following the reaction at 420 nm.

Determination of kinetic parameters of GH109_Pool2

Kinetic parameters were determinated by assaying 3 μg of enzyme, at the final concentration of 3.86 nM, in 50 mM sodium phosphate buffer pH 8.0, 1.5 mM NAD^+ on 4NP-β-Glc (0.05–5.0 mM), 4NP-β-GlcNAc (0.05–5.0 mM), 4NP-α-Glc (0.5–7.0 mM), and 4NP-α-GlcNAc (0.25–5.0 mM).

Thermal stability of GH109_Pool2

The thermostability of GH109_Pool2 (0.1 $mg{\cdot}mL^{-1}$) was evaluated by incubating the enzyme in PBS 1X (pH 7.3) at 75 and 85 °C up to 1 h. Aliquots (3 μg) were withdrawn at the indicated times and assayed at standard conditions. The residual activity was expressed as a percentage of the maximal enzymatic activity measured before the incubation at indicated temperatures.

Influence of different additives on GH109_Pool2

Assays (4.25 μg of GH109) were carried out at 85 °C by using 5 mM of 4NP-β-Glc as substrate, 2.5 mM NAD^+, 50 mM in sodium phosphate buffer pH 8.0 supplemented with different additives in a final volume of 500 μL. The enzyme activity was measured spectrophotometrically at 420 nm.

Acknowledgements

We want to thank the Beijing Genomics Institute, Hong Kong, Limited, BGI-Shenzhen, Shenzhen, China, for the deep sequencing and metagenomic analysis. This work was supported by a Short Term Mobility Program of the National Research Council of Italy and a grant from the Ministero dell'Università e della Ricerca Scientifica—Industrial Research Project 'Integrated agro-industrial chains with high energy efficiency for the development of eco-compatible processes of energy and biochemicals production from renewable sources and for the land valorization (Enerbio-Chem)' PON01_01966, funded in the frame of Operative National Programme Research and Competitiveness 2007–2013 D. D. Prot. n.01/Ric. 18.1.2010. We are grateful to Chiara Nobile and Marco Petruzziello at the Institute of Biosciences and BioResources (IBBR) from the National Research Council of Italy for administrative and technical assistance.

Conflicts of interest

The authors declare no conflict of interest.

Data accessibility

Research data pertaining to this article is located at figshare.com: https://doi.org/10.6084/m9.figshare.9983390.

Author contributions

ASt (Andrea Strazzulli), BCP, and MM conceived and designed the study; RI, RG, NC, LM, and CSdC performed experiments; ASt, FML, VL, and BH performed bioinformatic analyses; ASt, BCP, and MM collected and analyzed the data; PC and AS (Annalisa Santangelo) performed literature review; CF provided essential clones; ASt and MM wrote the manuscript.

References

1 Dumorne K, Cordova DC, Astorga-Elo M & Renganathan P (2017) Extremozymes: a potential source for industrial applications. *J Microbiol Biotechn* **27**, 649–659.
2 Kim H-S, Kim B-Y & Baik M-Y (2012) Application of Ultra High Pressure (UHP) in starch chemistry. *Crit Rev Food Sci Nutr* **52**, 123–141.
3 Singh G, Kaur K, Puri S & Sharma P (2015) Critical factors affecting laccase-mediated biobleaching of pulp in paper industry. *Appl Microbiol Biotechnol* **99**, 155–164.
4 Kumar V, Dangi AK & Shukla P (2018) Engineering thermostable microbial xylanases toward its industrial applications. *Mol Biotechnol* **60**, 226–235.
5 Limayem A & Ricke SC (2012) Lignocellulosic biomass for bioethanol production: current perspectives, potential issues and future prospects. *Prog Energ Combust* **38**, 449–467.
6 Valentine J, Clifton-Brown J, Hastings A, Robson P, Allison G & Smith P (2011) Food vs. fuel: the use of land for lignocellulosic 'Next Generation' energy crops that minimize competition with primary food production. *GCB Bioenergy* **4**, 1–19.
7 Akhtar J, Idris A & Abd. Aziz, R. (2014) Recent advances in production of succinic acid from lignocellulosic biomass. *Appl Microbiol Biotechnol* **98**, 987–1000.
8 Duwe A, Tippkötter N & Ulber R. Lignocellulose-Biorefinery: ethanol-focused, pp. 1–39. Springer, Berlin Heidelberg.
9 Bharathiraja B, Jayamuthunagai J, Sudharsanaa T, Bharghavi A, Praveenkumar R, Chakravarthy M &

Yuvaraj D (2017) Biobutanol - an impending biofuel for future: a review on upstream and downstream processing techniques. *Renew Sust Energ Rev* **68**, 788–807.
10 Yeoman CJ, Han Y, Dodd D, Schroeder CM, Mackie RI & Cann IKO (2010) Chapter 1 - Thermostable Enzymes as Biocatalysts in the Biofuel Industry. In Advances in Applied Microbiology (Jannasch HW eds), pp. 1–55. Academic Press, London.
11 Barnard D, Casanueva A, Tuffin M & Cowan D (2010) Extremophiles in biofuel synthesis. *Environ Technol* **31**, 871–88.
12 Sahm K, John P, Nacke H, Wemheuer B, Grote R, Daniel R & Antranikian G (2013) High abundance of heterotrophic prokaryotes in hydrothermal springs of the azores as revealed by a network of 16S rRNA gene-based methods. *Extremophiles* **17**, 649–662.
13 Jiang X & Takacs-Vesbach CD (2017) Microbial community analysis of pH 4 thermal Springs in Yellowstone National Park. *Extremophiles* **21**, 135–152.
14 Strazzulli A, Fusco S, Cobucci-Ponzano B, Moracci M & Contursi P (2017) Metagenomics of microbial and viral life in terrestrial geothermal environments. *Rev Environ Sci Bio* **16**, 425–454.
15 Antranikian G, Suleiman M, Schafers C, Adams MWW, Bartolucci S, Blamey JM, Birkeland NK, Bonch-Osmolovskaya E, da Costa MS, Cowan D *et al.* (2017) Diversity of bacteria and archaea from two shallow marine hydrothermal vents from Vulcano Island. *Extremophiles* **21**, 733–742.
16 Zaikova E, Benison KC, Mormile MR & Johnson SS (2018) Microbial communities and their predicted metabolic functions in a desiccating acid Salt Lake. *Extremophiles* **22**, 367–379.
17 Adesioye FA, Makhalanyane TP, Biely P & Cowan DA (2016) Phylogeny, classification and metagenomic bioprospecting of microbial acetyl xylan esterases. *Enzyme Microb Tech* **93–94**, 79–91.
18 Schröder C, Janzer V-A, Schirrmacher G, Claren J & Antranikian G (2017) Characterization of two novel heat-active α-galactosidases from thermophilic bacteria. *Extremophiles* **21**, 85–94.
19 Khan M & Sathya TA (2017) Extremozymes from metagenome: potential applications in food processing. *Crit Rev Food Sci Nutr* **58**, 2017–2025.
20 Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM & Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490–D495.
21 Zillig W, Stetter KO, Wunderl S, Schulz W, Priess H & Scholz I (1980) The *Sulfolobus*-"*Caldariella*" Group: taxonomy on the basis of the structure of DNA-dependent RNA polymerases. *Arch Microbiol* **125**, 259–269.
22 Segerer A, Neuner A, Kristjansson JK & Stetter KO (1986) *Acidianus infernus* Gen. Nov, Sp. Nov, and *Acidianus brierley*i Comb. Nov - facultatively aerobic, extremely acidophilic thermophilic sulfur-metabolizing archaebacteria. *Int J Syst Bacteriol* **36**, 559–564.
23 Huber G, Spinnler C, Gambacorta A & Stetter KO (1989) *Metallosphaera sedula* Gen. Nov and Sp. Nov represents a new genus of aerobic, metal-mobilizing, Thermoacidophilic Archaebacteria. *Syst Appl Microbiol* **12**, 38–47.
24 Huber R, Huber H & Stetter KO (2000) Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. *FEMS Microbiol Rev* **24**, 615–623.
25 Rodriguez-R LM, Gunturu S, Tiedje JM, Cole JR & Konstantinidis KT (2018) Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. *mSystems*, **3**, e00039-18.
26 Munson-McGee JH, Snyder JC & Young MJ (2018) Archaeal viruses from high-temperature environments. *Genes* **9**, 128. https://doi.org/10.3390/genes9030128
27 Krupovic M, Cvirkaite-Krupovic V, Iranzo J, Prangishvili D & Koonin EV (2018) Viruses of archaea: structural, functional, environmental and evolutionary genomics. *Virus Res* **244**, 181–193.
28 Langmead B & Salzberg SL (2012) Fast gapped-read alignment with bowtie 2. *Nat Methods* **9**, 357–U54.
29 Contursi P, Cannio R, Prato S, She QX, Rossi M & Bartolucci S (2007) Transcriptional analysis of the genetic element pSSVx: differential and temporal regulation of gene expression reveals correlation between transcription and replication. *J Bacteriol* **189**, 6339–6350.
30 Contursi P, Cannio R & She QX (2010) Transcription termination in the plasmid/virus hybrid pSSVx from *Sulfolobus islandicus*. *Extremophiles* **14**, 453–463.
31 Haring M, Vestergaard G, Rachel R, Chen L, Garrett RA & Prangishvili D (2005) Virology: independent virus development outside a host. *Nature* **436**, 1101–2.
32 Haring M, Peng X, Brugger K, Rachel R, Stetter KO, Garrett RA & Prangishvili D (2004) Morphology and genome organization of the virus PSV of the hyperthermophilic archaeal genera *Pyrobaculum* and *Thermoproteus*: a novel virus family, the Globuloviridae. *Virology* **323**, 233–242.
33 Wagner C, Reddy V, Asturias F, Khoshouei M, Johnson JE, Manrique P, Munson-McGee J, Baumeister W, Lawrence CM & Young MJ (2017) Isolation and characterization of *Metallosphaera turreted* icosahedral virus, a founding member of a new family of archaeal viruses. *J Virol* 91, e00925-17.
34 Lin HH & Liao YC (2016) Accurate binning of metagenomic contigs via automated clustering

sequences using information of genomic signatures and marker genes. *Sci Rep* **6**, 24175.
35 Parks DH, Imelfort M, Skennerton CT, Hugenholtz P & Tyson GW (2015) Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043–55.
36 van Wolferen M, Orell A & Albers SV (2018) Archaeal biofilm formation. *Nat Rev Microbiol* **16**, 699–713.
37 Kolbl D, Pignitter M, Somoza V, Schimak MP, Strbak O, Blazevic A & Milojevic T (2017) Exploring fingerprints of the extreme thermoacidophile *Metallosphaera sedula* grown on synthetic martian Regolith materials as the sole energy sources. *Front Microbiol* **8**, 1918.
38 Holmes DE, Risso C, Smith JA & Lovley DR (2012) Genome-scale analysis of anaerobic benzoate and phenol metabolism in the Hyperthermophilic Archaeon *Ferroglobus placidus*. *ISME J* **6**, 146–157.
39 Singh D, Mishra K & Ramanthan G (2015) Bioremediation of Nitroaromatic Compounds. In Wastewater Treatment Engineering (Samer M, ed), pp. 51–83, IntechOpen, London.
40 Copley SD (2009) Evolution of efficient pathways for degradation of anthropogenic chemicals. *Nat Chem Biol* **5**, 559–66.
41 Krzmarzick MJ, Taylor DK, Fu X & McCutchan AL (2018) Diversity and niche of archaea in bioremediation. *Archaea* **2018**, 3194108.
42 Valderrama JA, Durante-Rodriguez G, Blazquez B, Garcia JL, Carmona M & Diaz E (2012) Bacterial degradation of benzoate cross-regulation between aerobic and anaerobic pathways. *J Biol Chem* **287**, 10494–10508.
43 Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R *et al.* (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33**, 5691–5702.
44 Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP *et al.* (2008) A core gut microbiome in obese and lean twins. *Nature* **457**, 480.
45 CAZypedia Consortium (2018) Ten years of cazypedia: a living encyclopedia of carbohydrate-active enzymes. *Glycobiology* **28**, 3–8.
46 Nicolaus B, Manca MC, Romano I & Lama L (1993) Production of an exopolysaccharide from 2 thermophilic archaea belonging to the genus *Sulfolobus*. *FEMS Microbiol Lett* **109**, 203–206.
47 Horcajada C, Guinovart JJ, Fita I & Ferrer JC (2006) Crystal structure of an archaeal glycogen synthase - insights into oligomerization and substrate binding of eukaryotic glycogen synthases. *J Biol Chem* **281**, 2923–2931.
48 Meyer BH & Albers SV (2014) Aglb, catalyzing the oligosaccharyl transferase step of the archaeal N-glycosylation process, is essential in the thermoacidophilic *Crenarchaeon Sulfolobus acidocaldarius*. *Microbiologyopen*. **3**, 531–543.
49 Pouwels J, Moracci M, Cobucci-Ponzano B, Perugino G, van der Oost J, Kaper T, Lebbink JHG, de Vos WM, Ciaramella M & Rossi M (2000) Activity and stability of hyperthermophilic enzymes: a comparative study on two archaeal beta-glycosidases. *Extremophiles* **4**, 157–164.
50 Cobucci-Ponzano B, Moracci M, Di Lauro B, Ciaramella M, D'Avino R & Rossi M (2002) Ionic network at the C-terminus of the beta-Glycosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*: functional role in the quaternary structure thermal stabilization. *Proteins*. **48**, 98–106.
51 Morana A, Paris O, Maurelli L, Rossi M & Cannio R (2007) Gene cloning and expression in *Escherichia Coli* Of A Bi-functional beta-D-Xylosidase/Alpha-L-arabinosidase from *Sulfolobus solfataricus* involved in xylan degradation. *Extremophiles* **11**, 123–32.
52 Vecchio G, Parascandolo A, Allocca C, Ugolini C, Basolo F, Moracci M, Strazzulli A, Cobucci-Ponzano B, Laukkanen MO, Castellone MD & *et al.* (2017) Human alpha-L-fucosidase-1 attenuates the invasive properties of thyroid cancer. *Oncotarget* **8**, 27075–27092.
53 Limauro D, Cannio R, Fiorentino G, Rossi M & Bartolucci S (2001) Identification and molecular characterization of an endoglucanase gene, cels, from the extremely thermophilic archaeon *Sulfolobus solfataricus*. *Extremophiles* **5**, 213–219.
54 Maurelli L, Giovane A, Esposito A, Moracci M, Fiume I, Rossi M & Morana A (2008) Evidence that the xylanase activity from *Sulfolobus solfataricus O alpha* is encoded by the endoglucanase precursor gene (sso1354) and characterization of the associated cellulase activity. *Extremophiles* **12**, 689–700.
55 Fang TY, Hung XG, Shih TY & Tseng WC (2004) Characterization of the trehalosyl dextrin-forming enzyme from the thermophilic archaeon *Sulfolobus solfataricus ATCC 35092*. *Extremophiles* **8**, 335–343.
56 Woo EJ, Lee S, Cha H, Park JT, Yoon SM, Song HN & Park KH (2008) Structural insight into the bifunctional mechanism of the glycogen-debranching enzyme trex from the archaeon *Sulfolobus solfataricus*. *J Biol Chem* **283**, 28641–28648.
57 Kim MS, Park JT, Kim YW, Lee HS, Nyawira R, Shin HS, Park CS, Yoo SH, Kim YR, Moon TW *et al.* (2004) Properties of a novel thermostable glucoamylase from the hyperthermophilic archaeon *Sulfolobus solfataricus* in relation to starch processing. *Appl Environ Microbiol* **70**, 3933–3940.

58 Cobucci-Ponzano B, Conte F, Benelli D, Londei P, Flagiello A, Monti M, Pucci P, Rossi M & Moracci M (2006) The gene of an archaeal alpha-L-fucosidase is expressed by translational frameshifting. *Nucleic Acids Res* **34**, 4258–4268.
59 Ernst HA, Lo Leggio L, Willemoes M, Leonard G, Blum P & Larsen S (2006) Structure of the *Sulfolobus solfataricus* alpha-glucosidase: implications for domain conservation and substrate recognition in GH31. *J Mol Biol* **358**, 1106–1124.
60 Moracci M, Cobucci-Ponzano B, Trincone A, Fusco S, De Rosa M, van der Oost J, Sensen CW, Charlebois RL & Rossi M (2000) Identification and molecular characterization of the first alpha-xylosidase from an archaeon. *J Biol Chem* **275**, 22082–22089.
61 Trincone A, Cobucci-Ponzano B, Di Lauro B, Rossi M, Mitsuishi Y & Moracci M (2001) Enzymatic synthesis and hydrolysis of xylogluco-oligosaccharides using the first archaeal alpha-xylosidase from *Sulfolobus solfataricus*. *Extremophiles* **5**, 277–282.
62 Cobucci-Ponzano B, Conte F, Strazzulli A, Capasso C, Fiume I, Pocsfalvi G, Rossi M & Moracci M (2010) The molecular characterization of a novel GH38 alpha-mannosidase from the crenarchaeon *Sulfolobus solfataricus* revealed its ability in de-mannosylating glycoproteins. *Biochimie* **92**, 1895–907.
63 Worthington P, Hoang V, Perez-Pomares F & Blum P (2003) Targeted disruption of the alpha-amylase gene in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J Bacteriol* **185**, 482–488.
64 Cobucci-Ponzano B, Aurilia V, Riccio G, Henrissat B, Coutinho PM, Strazzulli A, Padula A, Corsaro MM, Pieretti G, Pocsfalvi G *et al*. (2010) A new archaeal beta-glycosidase from *Sulfolobus solfataricus*: seeding a novel retaining beta-glycan-specific glycoside hydrolase family along with the human non-lysosomal glucosylceramidase GBA2. *J Biol Chem* **285**, 20691–20703.
65 Ferrara MC, Cobucci-Ponzano B, Carpentieri A, Henrissat B, Rossi M, Arnoresano A & Moracci M (2014) The identification and molecular characterization of the first archaeal bifunctional exo-beta-glucosidase/*N*-acetyl-beta-glucosaminidase demonstrate that family GH116 is made of three functionally distinct subfamilies. *Biochim Biophys Acta* **1840**, 367–377.
66 Comfort DA, Chou CJ, Conners SB, VanFossen AL & Kelly RM (2008) Functional-genomics-based identification and characterization of open reading frames encoding alpha-glucoside-processing enzymes in the hyperthermophilic archaeon *Pyrococcus furiosus*. *Appl Environ Microb* **74**, 1281–1283.
67 Maki-Arvela P, Salmi T, Holmbom B, Willfor S & Murzin DY (2011) Synthesis of sugars by hydrolysis of hemicelluloses- a review. *Chem Rev* **111**, 5638–5666.
68 Neureiter M, Danner H, Thomasser C, Saidi B & Braun R (2002) Dilute-acid hydrolysis of sugarcane bagasse at varying conditions. *Appl Biochem Biotech* **98**, 49–58.
69 Marzialetti T, Olarte MBV, Sievers C, Hoskins TJC, Agrawal PK & Jones CW (2008) Dilute acid hydrolysis of loblolly pine: a comprehensive approach. *Ind Eng Chem Res* **47**, 7131–7140.
70 Kusema BT, Xu CL, Maki-Arvela P, Willfor S, Holmbom B, Salmi T & Murzin DY (2010) Kinetics of acid hydrolysis of arabinogalactans. *Int J Chem React Eng* **8**, ISSN (Online) 1542-6580
71 Xu CL, Pranovich A, Vahasalo L, Hemming J, Holmbom B, Schols HA & Willfor S (2008) Kinetics of acid hydrolysis of water-soluble spruce O-acetyl galactoglucomannans. *J Agr Food Chem* **56**, 2429–2435.
72 Liu QP, Sulzenbacher G, Yuan H, Bennett EP, Pietz G, Saunders K, Spence J, Nudelman E, Levery SB, White T *et al*. (2007) Bacterial glycosidases for the production of universal red blood cells. *Nat Biotechnol* **25**, 454–464.
73 Rahfeld P, Sim L, Moon H, Constantinescu I, Morgan-Lang C, Hallam SJ, Kizhakkedathu JN & Withers SG (2019) An enzymatic pathway in the human gut microbiome that converts A to universal O type blood. *Nat Microbiol* **4**, 1475–1485.
74 Aspeborg H, Coutinho PM, Wang Y, Brumer H & Henrissat B (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol*, **12**, 186.
75 Graham JE, Clark ME, Nadler DC, Huffer S, Chokhawala HA, Rowland SE, Blanch HW, Clark DS & Robb FT (2011) Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nat Commun* **2**, 375.
76 Zhang M, Jiang ZQ, Li LT & Katrolia P (2009) Biochemical characterization of a recombinant thermostable beta-mannosidase from *Thermotoga maritima* with transglycosidase activity. *J Mol Catal B-Enzym* **60**, 119–124.
77 Kim HW & Ishikawa K (2010) Structure of hyperthermophilic endocellulase from *Pyrococcus horikoshii*. *Proteins* **78**, 496–500.
78 Wang H, Squina F, Segato F, Mort A, Lee D, Pappan K & Prade R (2011) High-temperature enzymatic breakdown of cellulose. *Appl Environ Microbiol* **77**, 5199–5206.
79 Suleiman M, Schroder C, Klippel B, Schafers C, Kruger A & Antranikian G (2019) Extremely thermoactive archaeal endoglucanase from a shallow marine hydrothermal vent from Vulcano Island. *Appl Microbiol Biotechnol* **103**, 1267–1274.
80 Barras F, Bortoli-German I, Bauzan M, Rouvier J, Gey C, Heyraud A & Henrissat B (1992) Stereochemistry of the hydrolysis reaction catalyzed by endoglucanase Z from *Erwinia chrysanthemi*. *Febs Lett* **300**, 145–148.

81 Jenkins J, Lo Leggio L, Harris G & Pickersgill R (1995) Beta-Glucosidase, beta-galactosidase, Family A Cellulases, Family F xylanases and two barley glycanases form a superfamily of enzymes with 8-Fold beta/alpha architecture and with two conserved glutamates near the carboxy-terminal ends of beta-strands four and seven. *Febs Lett* **362**, 281–285.
82 Shi H, Huang Y, Zhang Y, Li W, Li X & Wang F (2013) High-Level expression of a novel thermostable and mannose-tolerant beta-mannosidase from *Thermotoga thermarum* DSM 5069 in *Escherichia coli*. *Bmc Biotechnol* **13**, 83.
83 Duffaud GD, McCutchen CM, Leduc P, Parker KN & Kelly RM (1997) Purification and characterization of extremely thermostable beta-mannanase, beta-mannosidase, and alpha-galactosidase from the hyperthermophilic eubacterium *Thermotoga neapolitana* 5068. *Appl Environ Microb* **63**, 169–177.
84 Zechel DL, He SM, Dupont C & Withers SG (1998) Identification of Glu-120 as the catalytic nucleophile in *Streptomyces lividans* endoglucanase CelB. *Biochem J* **336**, 139–145.
85 Tanaka M, Umemoto Y, Okamura H, Nakano D, Tamaru Y & Araki T (2009) Cloning and characterization of a beta-1,4-Mannanase 5C possessing a Family 27 carbohydrate-binding module from a marine bacterium, *Vibrio sp* Strain MA-138. *Biosci Biotech Bioch* **73**, 109–116.
86 Koshland DE (1953) Stereochemistry and the mechanism of enzymatic reactions. *Biol Rev* **28**, 416–436.
87 Yip VLY, Varrot A, Davies GJ, Rajan SS, Yang XJ, Thompson J, Anderson WF & Withers SG (2004) An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 beta-glycosidase from *Thermotoga maritima*. *J Am Chem Soc* **126**, 8354–8355.
88 Gupta N, Wu H & Terman JR (2016) Data presenting a modified bacterial expression vector for expressing and purifying Nus solubility-tagged proteins. *Data Brief* **8**, 1227–1231.
89 Yip VLY, Thompson J & Withers SG (2007) Mechanism of GlvA from *Bacillus subtilis*: a detailed kinetic analysis of a 6-phospho-alpha-glucosidase from glycoside hydrolase family 4. *Biochemistry* **46**, 9840–9852.
90 Bolger AM, Lohse M & Usadel B (2014) Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* **30**, 2114–2120.
91 Li D, Liu CM, Luo R, Sadakane K & Lam TW (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via Succinct de Bruijn Graph. *Bioinformatics* **31**, 1674–1676.
92 Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW & Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119
93 Fu LM, Niu BF, Zhu ZW, Wu ST & Li WZ (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152.
94 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K & Madden TL (2009) BLAST Plus : architecture and applications. *BMC Bioinformatics* **10**, 421.
95 Huson DH, Beier S, Flade I, Gorska A, El-Hadidi M, Mitra S, Ruscheweyh HJ & Tappu R (2016) MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *Plos Comput Biol* **12**, e1004957.
96 Buchfink B, Xie C & Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**, 59–60.
97 Cobucci-Ponzano B, Strazzulli A, Iacono R, Masturzo G, Giglio R, Rossi M & Moracci M (2015) Novel thermophilic hemicellulases for the conversion of lignocellulose for second generation biorefineries. *Enzyme Microb Tech* **78**, 63–73.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Taxonomy rarefaction analysis.

Fig. S2. Reads composition of the metagenome of solfatara Pisciarelli.

Fig. S3. Clustering of metagenomic contigs by MyCC.

Fig. S4. Percent distribution of the reads showing no match in NT on the clusters.

Fig. S5. Percentage identity distribution, against the NR database, of the putative ORFs obtained from the reads without match.

Fig. S6. BlastP score plot of the putative ORFs obtained from the reads without match against NR database.

Fig. S7. Taxonomic assignment at genus level of the ORFs obtained from the reads without match.

Fig. S8. COG analysis of the metagenome of solfatara Pisciarelli.

Fig. S9. Relative abundance of the sequences annotated to the KEGG Xenobiotic Biodegradation and Metabolism in Pool1 and Pool2.

Fig. S10. KEGG analysis of the ORFs obtained from the reads without match.

Fig. S11. SEED analysis of the ORFs obtained from the reads without match.

Fig. S12. Portion of the multialignment of GH5_Pool2 with other family members.

Fig. S13. Purification profile of GH5_Pool2.

Fig. S14. Molecular mass determination of GH5_Pool2.

Fig. S15. Activity of GH5_Pool2 on different concentration of 4NP-β-man.

Fig. S16. Thin Layer Chromatography of hydrolysis reactions with different substrates.

Fig. S17. Structures of the laminarin oligosaccharides and 63,64-α-galactosyl-mannopentaose.

Fig. S18. HPLC analysis of reaction products of GH5_Pool2 on laminaribiose.

Fig. S19. HPLC analysis of reaction products of GH5_Pool2 on laminartriose.

Fig. S20. HPLC analysis of reaction products of GH5_Pool2 on laminaritetraose.

Fig. S21. HPLC analysis of reaction products of GH5_Pool2 on laminaripentaose.

Fig. S22. Phylogenetic tree of the family GH109_Pool2.

Fig. S23. Purification profile of GH109_Pool2.

Fig. S24. Size-exclusion chromatography of GH109_Pool2.

Table S1. Analysis of the viral genomes in Pool1.

Table S2. Analysis of the viral genomes in Pool2.

Table S3. Cluster validation by CheckM.

Table S4. ORFs statistics from reads without match.

Table S5. Assignment of the ORFs from Pool1 and Pool2 to the COG functional classes.

Table S6. Assignment of the ORFs from Pool1 and Pool2 to the KEGG functional classes.

Table S7. Hemicellulosic composition of the plant species present in Pisciarelli solfatara.

Table S8. CAZymes identification in Pool1.

Table S9. CAZymes identification in Pool2.

Table S10. BlastP analysis of the ORFs from the contig_5232.

Table S11. Purification table of GH5_Pool2.

Table S12. Purification table of GH109_Pool2.

AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

ENZYMOLOGY AND PROTEIN ENGINEERING

GlcNAc De-*N*-Acetylase from the Hyperthermophilic Archaeon *Sulfolobus solfataricus*

Roberta Iacono,[a,b] Andrea Strazzulli,[a,c] Luisa Maurelli,[b] Nicola Curci,[a,b] Angela Casillo,[d] Maria Michela Corsaro,[d] Marco Moracci,[a,b,c] Beatrice Cobucci-Ponzano[b]

[a]Department of Biology, University of Naples Federico II, Naples, Italy
[b]Institute of Biosciences and BioResources, National Research Council of Italy, Naples, Italy
[c]Task Force on Microbiome Studies, University of Naples Federico II, Naples, Italy
[d]Department of Chemical Sciences, University of Naples Federico II, Naples, Italy

ABSTRACT *Sulfolobus solfataricus* is an aerobic crenarchaeal hyperthermophile with optimum growth at temperatures greater than 80°C and pH 2 to 4. Within the crenarchaeal group of *Sulfolobales*, *N*-acetylglucosamine (GlcNAc) has been shown to be a component of exopolysaccharides, forming their biofilms, and of the *N*-glycan decorating some proteins. The metabolism of GlcNAc is still poorly understood in *Archaea*, and one approach to gaining additional information is through the identification and functional characterization of carbohydrate active enzymes (CAZymes) involved in the modification of GlcNAc. The screening of *S. solfataricus* extracts allowed the detection of a novel α-*N*-acetylglucosaminidase (α-GlcNAcase) activity, which has never been identified in *Archaea*. Mass spectrometry analysis of the purified activity showed a protein encoded by the *sso2901* gene. Interestingly, the purified recombinant enzyme, which was characterized in detail, revealed a novel de-*N*-acetylase activity specific for GlcNAc and derivatives. Thus, assays to identify an α-GlcNAcase found a GlcNAc de-*N*-acetylase instead. The α-GlcNAcase activity observed in *S. solfataricus* extracts did occur when SSO2901 was used in combination with an α-glucosidase. Furthermore, the inspection of the genomic context and the preliminary characterization of a putative glycosyltransferase immediately upstream of *sso2901* (*sso2900*) suggest the involvement of these enzymes in the GlcNAc metabolism in *S. solfataricus*.

IMPORTANCE In this study, a preliminary screening of cellular extracts of *S. solfataricus* allowed the identification of an α-*N*-acetylglucosaminidase activity. However, the characterization of the corresponding recombinant enzyme revealed a novel GlcNAc de-*N*-acetylase, which, in cooperation with the α-glucosidase, catalyzed the hydrolysis of O-α-GlcNAc glycosides. In addition, we show that the product of a gene flanking the one encoding the de-*N*-acetylase is a putative glycosyltransferase, suggesting the involvement of the two enzymes in the metabolism of GlcNAc. The discovery and functional analysis of novel enzymatic activities involved in the modification of this essential sugar represent a powerful strategy to shed light on the physiology and metabolism of *Archaea*.

KEYWORDS *Archaea*, CAZymes, *N*-acetylglucosamine, de-*N*-acetylase

Archaeal glycobiology is a fascinating field of study, and recent discoveries shed light on a significant structural and functional diversity of carbohydrates in this domain of life. Besides being carbon and energy sources, glycans are essential parts of the extracellular matrix of the archaeal cell (1), the cell wall, and the N-linked glycosylation of proteins (2). The extracellular matrix surrounding most thermophiles and halophiles (1) is composed by exopolysaccharides (EPS). In *Sulfolobales*, EPS production

Citation Iacono R, Strazzulli A, Maurelli L, Curci N, Casillo A, Corsaro MM, Moracci M, Cobucci-Ponzano B. 2019. GlcNAc de-*N*-acetylase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. Appl Environ Microbiol 85:e01879-18. https://doi.org/10.1128/AEM.01879-18.

Editor Robert M. Kelly, North Carolina State University

Address correspondence to Marco Moracci, marco.moracci@unina.it, or Beatrice Cobucci-Ponzano, beatrice.cobucciponzano@ibbr.cnr.it.

Received 31 July 2018
Accepted 4 November 2018
Accepted manuscript posted online 16 November 2018
Published 9 January 2019

is used to adapt to rapid variations of temperature, pH, and geochemical conditions and for cell adhesion, as well as establishing a carbohydrate reservoir to be exploited either as an energy source or during biofilm formation (3). The majority of characterized *Archaea* have a proteinaceous cell wall, consisting of a regularly structured two-dimensional array based either on a single protein species (S-layer) or a limited number of proteins (S-layer glycoprotein). This structure provides the basis for the ability of many archaeal species to thrive in extreme habitats (4). Some S-layer glycoproteins are modified by N-linked glycosylation, a universal process in *Archaea* that is important in motility, cellular adhesion, cell-cell communication, biofilm formation, and the maintenance of the cellular shape (5–9) and is essential for the viability of the crenarchaeal major phylum (10).

N-Acetylglucosamine (GlcNAc) has been identified in EPS and biofilms (11, 12), in glycans N-linked to proteins (5, 10, 13, 14), and in a biosynthetic intermediate of the glycosylphosphatidylinositol (GPI) anchor (15). More specifically, within the crenarchaeal group of *Sulfolobales*, GlcNAc has been shown to be a component of the EPS, forming their biofilms (3, 16), and of the N-glycan, decorating the cytochrome $b_{558/566}$ (17) and an ABC transporter of *S. solfataricus* (18). In *Archaea*, the metabolism of GlcNAc is still poorly understood, but this can be elucidated through the identification and functional characterization of carbohydrate active enzymes (CAZymes) involved in its modification.

S. solfataricus (strain P2) is an aerobic chemoorganoheterotroph crenarchaeon with optimal growth at 80°C (range, 60 to 92°C) and pH 2.0 to 4.0 (19). It is able to grow heterotrophically on various carbon and energy sources, and this is reflected by its noticeably high number of genes encoding glycoside hydrolases (GHs) (20–23). For this reason, it represents a source of new enzymes with unique properties in term of biotechnological applications (24, 25), thus making its study of high interest. In addition, the abundance of N-glycan chains in *Sulfolobus* spp. (18, 26, 27) suggests that part of the CAZyme repertoire in this organism could be involved in the recycling of carbohydrates, similarly to what occurs in *Bacteria* (28). This has been already proposed for *S. solfataricus*, since it has 27 potential GHs from 14 different families (CAZy classification, June 2018). Of these GHs, a β-glycosidase from GH1 (25) and an α-mannosidase from GH38 (21) might be involved *in vivo* in the modulation of the sugar composition of the EPS and in the demannosylation of the glycan tree of the extracellular glycoproteins (29). Moreover, we recently identified in *S. solfataricus* P2 a novel β-*N*-acetylglucosaminidase that may be involved in the turnover of the *N*,*N*-diacetylchitobiose of glycoproteins (20). Nevertheless, no α-*N*-acetylglucosaminidase activity has been identified in *Archaea*. In this study, we show that cellular extracts of *S. solfataricus* P2 contain α-*N*-acetylglucosaminidase activity. However, the purification of the native enzyme, as well as the detailed characterization by steady-state kinetics of the recombinant enzyme and chromatographic and nuclear magnetic resonance (NMR) analysis of the reaction products, revealed a novel and unexpected GlcNAc de-*N*-acetylase, which, in cooperation with the α-glucosidase MalA, catalyzed the hydrolysis of *O*-α-GlcNAc glycosides. The characterization of this enzyme and the association of its gene with the one encoding a putative glycosyltransferase (GT) will help to elucidate the GlcNAc metabolism in *Crenarchaeota*.

RESULTS

Purification of a putative α-*N*-acetylglucosaminidase activity from *S. solfataricus* P2. Free cellular extracts (FCE) were prepared from a cell culture of *S. solfataricus* P2 stopped in the early stationary phase. Assays at 65°C on 4Np- and 2Np-α-GlcNAc determined an increment in the absorbance at 420 nm indicating a putative *N*-acetylglucosaminidase activity. The assay allowed the purification of a protein through three chromatographic steps. Although the purification was performed from 9.5 g of *S. solfataricus* culture, the procedure yielded only tiny amounts of pure protein, detectable by SDS-PAGE with SYPRO Orange staining (Fig. 1A).

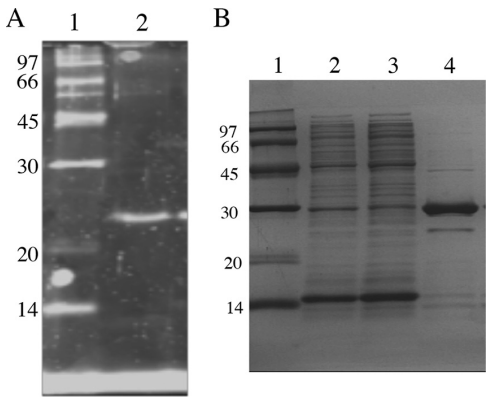


FIG 1 Purification of native and recombinant SSO2901. (A) SDS-PAGE of the sample from *S. solfataricus* extracts stained with Sypro Orange. Lane 1, molecular weight markers (kDa); lane 2, sample corresponding to the active fraction after gel filtration. (B) SDS-PAGE of rSSO2901 purification steps. Lane 1, molecular weight markers (kDa); lane 2, cell-free extract; lane 3, IMAC flowthrough; lane 4, IMAC elution.

Identification of the enzyme purified from *S. solfataricus*. To identify the open reading frame (ORF) encoding the α-*N*-acetylglucosaminidase activity, the protein band reported in Fig. 1A, lane 2, was excised from the gel and analyzed by tandem mass spectrometry (MS/MS), allowing the clear identification of a hypothetical protein codified by the ORF *sso2901* (matches 19, score 392) (Mass Spectrometry Service/CEINGE, Naples, Italy). SSO2901 showed a predicted molecular mass of 22,342 Da, perfectly matching the protein migration on SDS-PAGE (Fig. 1A, lane 2). BLASTp analyses revealed that SSO2901 encodes a putative *N*-acetyl-glucosaminyl-phosphatidylinositol de-*N*-acetylase (GlcNAc-PI de-*N*-acetylase). It is worth noting that none of the proteins identified by liquid chromatography-MS/MS (matches 2, score 33) were similar to putative α-*N*-acetylglucosaminidases. Therefore, either the identified SSO2901 showed both de-*N*-acetylase and α-*N*-acetylglucosaminidase activities or the hydrolysis of 4Np-α-GlcNAc, observed by monitoring the absorbance at 420 nm, occurred through a different mechanism. To test these hypotheses, the native SSO2901 was assayed on 4Np-α-GlcNAc and GlcNAc, and the reaction mixtures were analyzed by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) (Fig. 2). Analysis of the reaction of the purified SSO2901 on 4Np-α-GlcNAc (Fig. 2A) revealed the production of an unknown product in the reaction mixture at 5 min, while GlcNAc was present both in the blank and reaction mixture, possibly due to the spontaneous hydrolysis of the substrate. In contrast, the analysis of the reaction on GlcNAc clearly revealed the production of a peak corresponding to glucosamine (GlcN) (Fig. 2B), suggesting that SSO2901 has de-*N*-acetylase but not α-*N*-acetylglucosaminidase activity.

Sequence analysis of SSO2901. According to BLASTp alignments, SSO2901 encodes a putative GlcNAc-PI de-*N*-acetylase, belonging to the LmbE-like superfamily. This superfamily includes *N*-acetyl-glucosaminyl-phosphatidylinositol (GlcNAc-PI) de-*N*-acetylases, 1-D-myo-inosityl-2-acetamido-2-deoxy-α-D-glucopyranoside (GlcNAc-Ins) de-*N*-acetylase (30), *N*,*N*-diacetylchitobiose de-*N*-acetylase (31), and antibiotic de-*N*-acetylase (32).

The GlcNAc-Ins and *N*,*N*-diacetylchitobiose de-*N*-acetylases are classified into the family CE14 of the carbohydrate esterases in the CAZy database. The multialignments (obtained by using ClustalO) of the SSO2901 sequence with those from different carbohydrate esterase (CE) families clearly indicate that SSO2901 belongs to the CE14 family (data not shown). Both the multialignments and the phylogenetic tree (see Fig. S1 in the supplemental material) obtained comparing CE14 sequences with more closely related (CE11 and CE12) or more phylogenetically distant (CE6) CE families, confirm that SSO2901 shares a high sequence similarity with the other members of

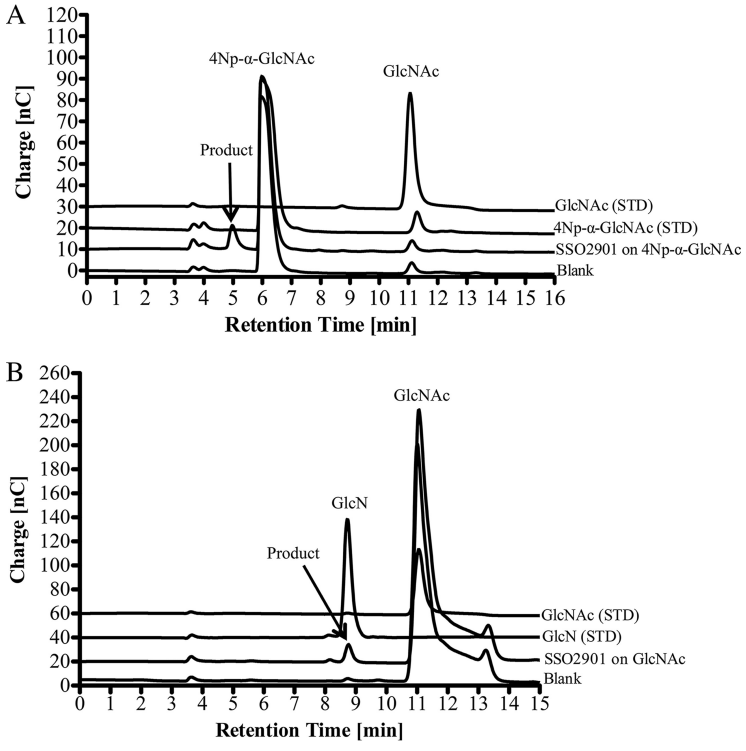


FIG 2 (A and B) HPAEC-PAD profiles of the enzymatic assays of SSO2901 on 4Np-α-GlcNAc (A) and GlcNAc (B) and of the relative blank mixtures. In the blank mixtures, a 5 mM concentration of the substrate was incubated without enzyme under the same conditions (50 mM phosphate buffer [pH 6.5] for 16 h at 65°C in 0.2 ml). The arrows indicate the products of the enzymatic activity.

CE14. In fact, SSO2901 showed 100% identity with Ssol_0697 from *S. solfataricus* 98/2, which is already classified into this family, together with the homologous sequences from *S. islandicus* REY15A (ADX84495.1), *S. islandicus* LAL14/1 (AGJ61906.1), *S. islandicus* L.S.2.15 (ACP34631.1), and *Metallosphaera cuprina* Ar-4 (AEB95994.1).

Cloning, expression, and characterization of recombinant SSO2901. To biochemically characterize SSO2901, its gene was cloned in the expression vector pET101/d-TOPO in-frame with a C-terminal His tag. The purification of the recombinant enzyme (rSSO2901) by immobilized metal affinity chromatography (IMAC) led to a >90% pure enzyme preparation with a final *Escherichia coli* yield of 6 mg liter^{-1} (Fig. 1B). The molecular mass of the recombinant enzyme in solution was 21.4 ± 1.48 kDa, which is comparable to the one observed for the denatured monomer, indicating that rSSO2901 is a monomer in its native form. The recombinant rSSO2901 (3 μg) was assayed on 4Np-α- and β-GlcNAc (Fig. 3A and B, respectively) under the same conditions used for the native enzyme (50 mM phosphate [pH 6.5], 65°C for 16 h). The analysis of the reaction mixtures by HPAEC-PAD confirmed the activity on 4Np-α-GlcNAc (compare Fig. 3A and Fig. 2A) and revealed that the enzyme is also active on 4Np-β-GlcNAc (Fig. 3B).

To unequivocally identify the products of rSSO2901, the reaction mixture of the enzyme assayed on 2Np-α-GlcNAc in 50 mM phosphate (pH 6.5) at 65°C for 16 h was analyzed by thin-layer chromatography (TLC) (Fig. S2) and NMR (Table 1 and Fig. S3 and S4). The ^{1}H,^{13}C DEPT-HSQC experiment (Fig. S3) of the enzymatic reaction revealed the presence of signals attributable to three different products, namely A, B, and C, in the relative amounts of 1:0.5:0.3. None of these signals was in agreement with those of 2Np-α-GlcNAc (Table 1), the chemical shifts of which were obtained from a separate

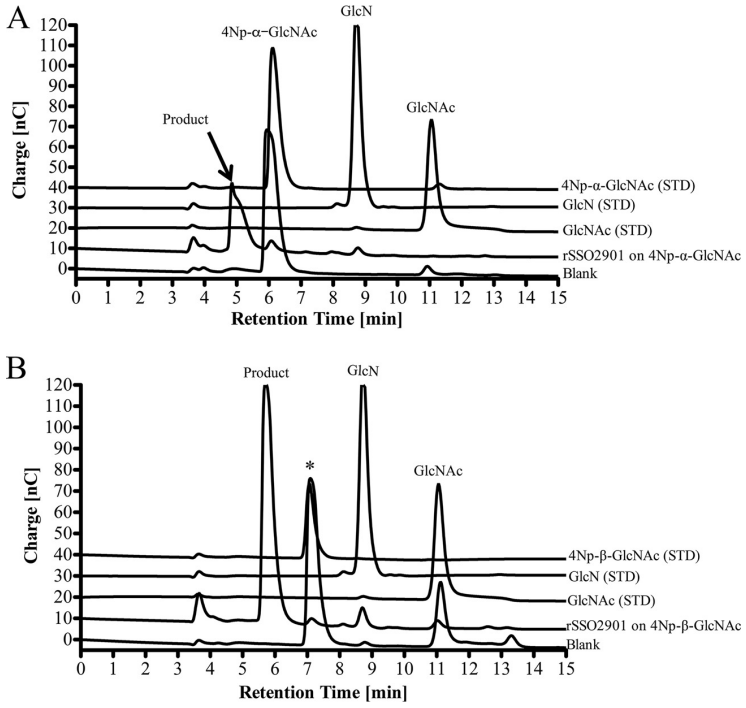


FIG 3 (A and B) HPAEC-PAD profiles of the enzymatic assays of rSSO2901 on 4Np-a-GlcNAc (A) and 4Np-b-GlcNAc (B) and of the relative blank mixtures. In the blanks, a 5 mM concentration of the substrate was incubated without enzyme under the same conditions (50 mM phosphate buffer [pH 6.5] for 16 h at 65°C in 0.2 ml) of the relative enzymatic assay. The arrows indicate the products obtained by enzymatic activity. The asterisk indicates the 4Np-b-GlcNAc peak.

^{1}H,^{13}C DEPT-HSQC experiment (Fig. S4). The most intense H-1/C-1 signals at δ 5.74/97.5 ppm (product A, Table 1, Fig. S3) was attributable to 2Np-α-GlcNH$_2$, as a result of its H-2/C-2 chemical shifts at δ 3.02/55.8 ppm. In particular, the shift of the H-2 signal from δ 4.02 ppm in the 2 Np-α-GlcNAc (Table 1) to δ 3.02 ppm in the product A of the reaction of rSSO2901 clearly indicated the absence of the amide group at the N-2 position (33, 34). The disappearance of the –CH$_3$ signals of acetyl group at δ 1.94/23.0 ppm (Table 1) confirmed the hydrolysis of the amide linkage. The less intense anomeric signals at δ 5.25/91.8 and 4.80/97.6 ppm were attributable to α and β reducing glucosamine residues, respectively (products B and C, Fig. 4, Table 1), based on their carbon chemical shifts. Since in both residues the H-2 signals were found at δ 3.65 and 3.61 ppm, respectively, it was deduced that the nitrogen atoms were substituted. An in-depth study of all two-dimensional NMR experiments revealed that products B and C were the 2-[*N*-(2-nitrophenyl)amino]-2-deoxy-D-glucopyranose [α-GlcNH(2Np)] α and β anomers, respectively. In fact, the proton and carbon chemical shift values of products B and C were in good agreement with those published for similar compounds, which are reported to be yellow (33). The finding of the 2-[*N*-(2-nitrophenyl)amino]-2-deoxy-D-glucopyranose is probably due to an enzymatic intramolecular transfer of the 2-nitrophenyl aglycone from the anomeric position to the deacetylated nitrogen atom, in a nucleophilic aromatic substitution through the formation of a Meisenheimer-like complex (34). The NMR analysis clearly demonstrates that SSO2901 is a de-*N*-acetylase active on GlcNAc derivatives. In addition, the formation of the yellow secondary products, revealed by NMR analysis, clarifies the results obtained from the assays of extracts of *S. solfataricus* P2 on the chromogenic substrates 2Np-α- and 4Np-α-GlcNAc. On the basis of these results, the subsequent characterization has been performed on GlcNAc using fluorescamine assay (FSA).

TABLE 1 ^{1}H and ^{13}C chemical shifts in D_2O at 600 MHz of the products of the reaction of rSSO2901

Residue	H-1, C-1	H-2, C-2	H-3, C-3	H-4, C-4	H-5, C-5	$H\text{-}6_{a,b}$, C-6	*N*-Acetyl H/C	Phenyl H-3′/C-3′	Phenyl H-4′/C-4′	Phenyl H-5′/C-5′	Phenyl H-6′/C-6′
2Np-α-GlcNAc	5.74, 97.5	4.02, 54.8	3.86, 71.9	3.49, 70.9	3.67, 74.6	3.66, 61.5	1.94, 23.0	7.85, 127.0	7.14, 124.0	7.56, 136.3	7.33, 119.0
2Np-α-GlcNH$_2$ (product A)	5.80, 98.4	3.02, 55.8	3.75, 74.0	3.44, 70.7	3.66, 75.0	3.66, 61.3		7.88, 127.0	7.14, 123.8	7.57, 136.6	7.40, 138.5
α-GlcNH(2Np) (product B)	5.25, 91.8	3.85, 57.8	3.78, 74.0	3.49, 70.8	3.78, 74.0	3.72, 3.78, 61.7		8.06, 128.0	6.66, 118	7.10, 116.0	7.44, 138.6
β-GlcNH(2Np) (product C)	4.79, 97.6	3.61, 61.4	3.66, 74.6	3.66, 74.6	3.44, 77.0	3.83, 62.0		8.06, 128.0	6.66, 118	7.10, 116.0	7.44, 138.6

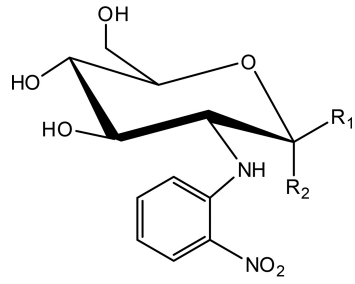


FIG 4 Structures of products B (α anomer; R_1 = H, R_2 = OH) and C (β anomer; R_1 = OH, R_2 = H).

To determine the pH optimum, rSSO2901 was assayed in 50 mM different buffers in the range of pH 5.0 to 10.0 on 20 mM GlcNAc using FSA. The enzyme is optimally active in HEPES buffer at pH 8.5 (Fig. 5A). The temperature profile of the enzyme was assessed in the range of 40 to 85°C on 20 mM GlcNAc in 50 mM buffer HEPES (pH 8.5); the maximal activity was reached at 70°C (Fig. 5B). In addition, rSSO2901 showed a remarkable stability when incubated for 120 min at 60 and 70°C, maintaining 100 and 60% of its activity, respectively, while it was completely inactivated after 5 min at 80°C (Fig. 5C). To study the effect of different metal ions and on rSSO2901, metal ions or EDTA were individually added to the reaction solution to a final concentration of 1 mM, as reported for other de-*N*-acetylases (35). De-*N*-acetylase activity was determined on 20 mM GlcNAc, in 50 mM HEPES (pH 8.5) at 70°C (Fig. 5D). As shown in Fig. 5D, rSSO2901 showed about 50% inactivation by Mn^{2+}, Zn^{2+}, and EDTA. On the basis of these results, the following characterization was performed in 50 mM HEPES buffer at pH 8.5 and at 70°C (rSSO2901 standard conditions).

To determine the substrate specificity, rSSO2901 was assayed under standard conditions on various *N*-acetyl-glycoside derivatives and analyzed by FSA or HPAEC-PAD. Among the tested substrates, rSSO2901 was active only on GlcNAc and its

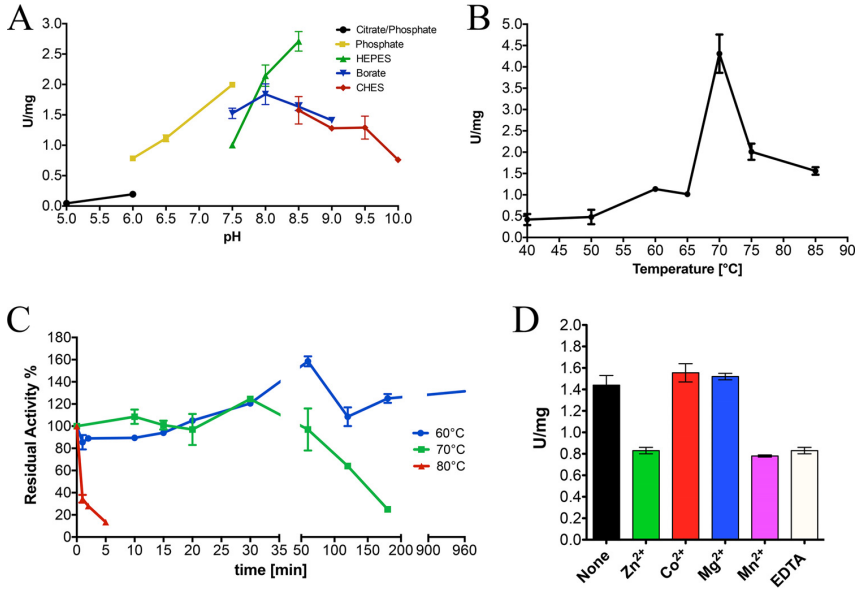


FIG 5 Enzymatic characterization of rSSO2901. (A) pH dependence of rSSO2901; (B) temperature dependence; (C) thermostability of rSSO2901. (D) Enzymatic assays were performed in the presence of different metal ions.

TABLE 2 Steady-state kinetic constants of rSSO2901 on different substrates

Substrate	Mean ± SD			
	V_{max} (U/mg)	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} mM^{-1})
2Np-α-GlcNAc	1.26 ± 0.30	6.55 ± 4.00	0.45 ± 0.11	0.07
MU-α-GlcNAc	0.42 ± 0.04	0.82 ± 0.24	0.15 ± 0.02	0.18
GlcNAc	5.44 ± 0.55	22.01 ± 4.91	1.94 ± 0.19	0.09

aryl-derivatives. No activity was found on GlcNAc-1P, GlcNAc-6S, GalNAc, ManNAc, UDP-GlcNAc, or *N*,*N*-diacetylchitobiose (not shown).

The initial aim of this work was to identify a α-*N*-acetylglucosaminidase (α-GlcNAcase) activity in *S. solfataricus*. However, the enzymatic assays on 2Np- and 4Np-α-GlcNAc could not demonstrate whether α-GlcNAcase activity was really absent in *S. solfataricus* or was hidden by the production of the chromogenic subproduct [α-GlcNH(2Np)] of SSO2901. Thus, the *S. solfataricus* cellular extracts were tested on 4-methylumbelliferyl-α-GlcNAc (4MU-α-GlcNAc) since the de-*N*-acetylation of this substrate did not produce a fluorescent product. After 16 h of incubation, an increase of fluorescence intensity was measured (117 arbitrary units) compared to blank mixture, suggesting that *S. solfataricus* is able to metabolize α-*N*-acetyl-glucosaminides. The α-glucosidase (MalA), encoded by *sso3051* (36), and the α-xylosidase (XylS), encoded by *sso3022* (37), which are both available α-glycosidases characterized in *S. solfataricus* and belonging to the family GH31, were singularly tested on 4MU-α-GlcNAc. Both enzymes did not show activity on 4MU-α-GlcNAc. Thus, we wondered whether the α-*N*-acetylglucosaminidase activity could be due to the cooperative action of two or more enzymes in the cellular extracts. To test this hypothesis, rSSO2901 was assayed on 4MU-α-GlcNAc in the presence of MalA or XylS, since the latter α-glycosidases could be active on the 4MU-GlcN produced by rSSO2901. As shown in Fig. S5, an increase in fluorescence emission was observed when SSO2901 acted in cooperation with MalA.

The steady-state kinetic constants of rSSO2901 on GlcNAc and 4MU-α-GlcNAc are reported in Table 2. The highest specificity constant was observed on 4MU-α-GlcNAc, followed by GlcNAc.

Analysis of the genomic environment of SSO2901. To shed light on the possible role of SSO2901 *in vivo*, we analyzed its genomic environment, revealing the presence of the ORF *sso2900*, immediately upstream of *sso2901* under the same transcriptional orientation. The *sso2900* gene encodes a putative glycosyl transferase belonging to GT4, a family grouping more than 20 different enzymatic activities, including the UDP-GlcNAc transferase. In an effort to determine the enzymatic activity of SSO2900, its gene was cloned and expressed in *E. coli*. The recombinant enzyme was purified using IMAC, obtaining >90% pure enzyme with a yield of 2 mg $liter^{-1}$ of culture (Fig. S6). Recombinant SSO2900 (30 μg) was incubated for 16 h in 50 mM Tris-HCl buffer (pH 7.5) at 65°C with 10 mM UDP-GlcNAc and with or without myoinositol as an acceptor. The reaction mixtures, analyzed by HPAEC-PAD, showed that rSSO2900 hydrolyzed UDP-GlcNAc, regardless of the presence of the acceptor, whereas transglycosylation peaks were not observed (Fig. 6). This may suggest that rSSO2900 recognized UDP-GlcNAc as a donor, but, in the absence of a suitable acceptor, promoted only substrate hydrolysis with low efficiency, as observed for other GTs (38).

DISCUSSION

In *Sulfolobales*, *N*-acetylglucosamine has been identified among the monosaccharides of N-glycans of proteins, EPS, and as a biosynthetic intermediate of the glycosylphosphatidylinositol (GPI) anchor (15). However, the enzymes involved in the turnover of these glycoconjugates are still unknown. Searches in *S. solfataricus* P2 genome sequence led to the discovery and characterization of a GH38 α-mannosidase and a GH116 bifunctional β-glucosidase/xylosidase (SSO1353), possibly involved in the protein N-glycosylation and EPS turnover (22, 29). Recently, GH116 bifunctional exo-β-glucosidase/exo-β-*N*-acetylglucosaminidase (SSO3039), likely involved in the recycling

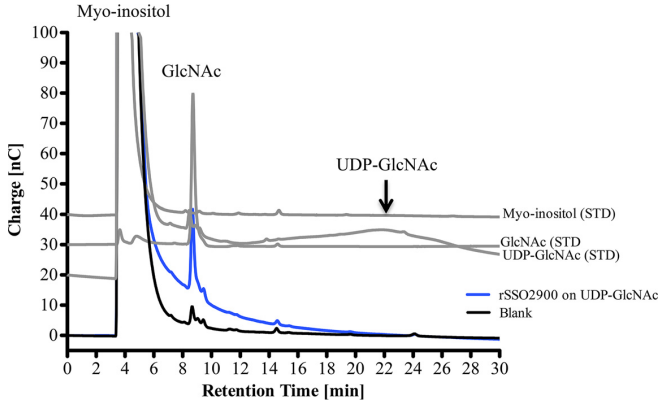


FIG 6 HPAEC-PAD analysis of rSSO2900 reaction on UDP-GlcNAc and myoinositol. In the blanks, 10 mM concentrations of both UDP-GlcNAc and myoinositol were incubated without enzyme under the same conditions (50 mM Tris-HCl [pH 7.5] for 16 h at 65°C in 0.2 ml) in the relative enzymatic assay. Blue lines indicate chromatograms of the reaction. Gray lines indicate chromatograms of the standards.

of *N*-acetylglucosamine in *S. solfataricus*, has been identified and characterized (20). In contrast, α-*N*-acetylglucosaminidase activities (EC 3.2.1.50), currently classified only in GH89, are unknown in *Archaea*. The activity screening for a α-*N*-acetylglucosaminidase in *S. solfataricus* P2, allowed us to identify the enzyme SSO2901. Surprisingly, this activity was demonstrated to be, instead, a GlcNAc de-*N*-acetylase. The enzymatic characterization of rSSO2901 revealed that it is active on GlcNAc, MU-α-, 2 Np-α-, 4 Np-α-, and β-GlcNAc, indicating that both anomers were tolerated in the active site. This was previously observed also for the de-*N*-acetylase from *Trypanosoma brucei* that was active on both GlcNAc-α- and GlcNAc-β-PI (39).

SSO2901 shows sequence similarity to a putative *N*-acetylglucosaminyl-phosphatidylinositol (GlcNAc-PI) de-*N*-acetylase from *Sulfolobus islandicus* belonging to the LmbE-like superfamily, which includes metal-dependent enzymes (40). All LmbE-like superfamily members possess the conserved His-X-Asp-Asp sequence (40), where the catalytic Zn^{2+} ion in the active site is pentacoordinated by the imidazolium side chain of His. The positively charged Zn^{2+} plays an essential role in stabilizing the reaction intermediate during catalysis (41). It has been reported that mononuclear zinc enzymes are inhibited by EDTA and Zn^{2+} (40, 42), as demonstrated for the rat GlcNAc-PI de-*N*-acetylase (43), for BC1534 and BC3461 from *B. cereus* (44), and for the de-*N*-acetylase LpxC from *E. coli* (45). rSSO2901 shows the conserved His-X-Asp-Asp motif, like the LmbE-like superfamily members (40) (Fig. S7), and it is partially inactivated by EDTA and Zn^{2+}, suggesting the presence of a catalytic metal ion in the active site of the enzyme.

The LmbE-like superfamily includes de-*N*-acetylases involved in four metabolic pathways: (i) the 1-D-myo-inosityl-2-acetamido-2-deoxy-α-D-glucopyranoside (GlcNAc-Ins) de-*N*-acetylase from *Mycobacterium tuberculosis* (EC 3.5.1.89) in mycothiol biosynthesis (30); (ii) the *N*,*N*-diacetylchitobiose de-*N*-acetylase from the archaeon *Thermococcus kodakaraensis* KOD1 (EC 3.5.1.-) in chitin degradation (31); (iii) the antibiotic de-*N*-acetylase from *Actinoplanes teichomyceticus* (Orf2) in the modification of lipoglycopeptidic antibiotic teicoplanin (32); and (iv) *N*-GlcNAc-PI de-*N*-acetylases (EC 3.5.1.89) from mammals, yeast, and protozoa in the biosynthesis of the GPI anchor (46–48). Antibiotic and GlcNAc-PI de-*N*-acetylases are not yet classified in CAZy (49). On the contrary, GlcNAc-Ins and *N*,*N*-diacetylchitobiose de-*N*-acetylases are classified into the CE14 family of carbohydrate esterases in the CAZy database. The pathways of mycothiol and bacillithiol biosynthesis are absent in *S. solfataricus*, which produces glutathione for the maintenance of redox homeostasis (50). This suggests that SSO2901 is not functionally correlated to the bacterial CE14 de-*N*-acetylases from *Bacillus cereus*,

Mycobacterium tuberculosis, and *M. smegmatis* (51–53). In addition, SSO2901 does not seem to be correlated to the archaeal CE14 *N*,*N*-diacetylchitobiose deacetylases from *Haloferax mediterranei*, *Pyrococcus furiosus*, *P. horikoshii*, and *Thermococcus kodakarensis* involved in chitin degradation (31, 54) since, to our knowledge, *S. solfataricus* is not able to grow on chitin, and no hypothetical chitinases were found in its genome. However, *N*,*N*-diacetylchitobiose is a key component of the prominent glycan structure, which is characteristic of a number of different surface-exposed proteins of *S. solfataricus* (18). Thus, although rSSO2901 is not active on *N*,*N*-diacetylchitobiose, its involvement in the turnover of *N*,*N*-diacetylchitobiose in N-glycans by deacetylating GlcNAc released by the β-*N*-acetylglucosaminidase SSO3039 (20) cannot be ruled out.

SSO2901 shows similarity with GlcNAc-PI deacetylases of the LmbE-like superfamily involved in the biosynthesis of the GPI anchor, a posttranslational modification composed of complex glycolipids covalently linked to the C terminus of proteins (55) and providing a stable anchor to the membrane (56, 57). The biosynthetic pathway begins with the transfer of *N*-acetylglucosamine from the UDP-*N*-acetylglucosamine (UDP-GlcNAc) donor to PI by glycosyltranferase. The resulting GlcNAc-PI is deacetylated to GlcN-PI (58). In *Archaea*, the presence of the GPI anchor has been demonstrated by Kobayashi et al. in *S. acidocaldarius*, but the enzymes involved in its biosynthesis are still unknown (15). In *S. solfataricus* P2, immediately upstream of *sso2901*, the *sso2900* gene encodes a putative GT4 glycosyltransferase. The comparison with *S. acidocaldarius* revealed that SSO2900 and SSO2901 are similar to the uncharacterized ORFs SUSAZ_08645 (putative GT4) and SUSAZ_05460 (putative CE14), sharing 28 and 26% identity, respectively. The preliminary enzymatic characterization of rSSO2900 revealed that it recognizes UDP-GlcNAc as a donor, suggesting its putative GT function and a possible functional correlation of the two enzymes *in vivo*.

To our knowledge, the presence of α-*N*-acetylglucosaminides in *Sulfolobus* sp. has not been demonstrated. However, the stereochemistry of the linkage between the monosaccharidic components in EPS and biofilms is still unknown. In bacterial biofilms, *N*-acetylglucosamine has been found in both α- and β-bonds (59, 60), while in *Eukarya*, the glucosaminoglucan heparan sulfate is composed by the repeating disaccharide unit of an uronic acid and *N*-acetylglucosamine [GlcAβ or IdoA2Xα-1-4-GlcNY3X6X-α1]$_n$ (61). Although it is premature to suggest the function of SSO2901 *in vivo*, this study shows that *S. solfataricus* is able to hydrolyze α-GlcNAc derivatives, and the α-*N*-acetylglucosaminides, when present, can be hydrolyzed through the concerted action of SSO2901 and MalA. Further analyses, including *in vivo* studies, could shed light on the many biological pathways in which GlcNAc-modifying enzymes are involved.

MATERIALS AND METHODS

All commercially available substrates (2Np- and 4Np-α-GlcNAc, MU-α-GlcNAc, 4Np-β-GlcNAc, UDP-GlcNAc, GlcNAc-1P, GlcNAc-6S, GalNAc, ManNAc, and *N*,*N*-diacetylchitobiose) were purchased from Sigma-Aldrich and Carbosynth unless otherwise designated.

***Archaea* and bacterial strains.** The strains used in this work were as follows: *Sulfolobus solfataricus* P2 (19), *E. coli* TOP10 [F$^-$ *mcrA* Δ(*mrr-hsdRMS-mcrBC*)ϕ80*lacZ*ΔM15 Δ*lacX74 recA1 araD139* Δ(*ara leu*)*7697 galU galK rpsL* (Str^r) *endA1 nupG*], and *E. coli* BL21 Star(DE3) [F$^-$ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm rne131*(DE3); Invitrogen].

Culture media. *S. solfataricus* Brock medium (57) was adjusted to pH 3.5 with sulfuric acid and supplemented with yeast extract, sucrose, and Casamino Acids (0.1% each) as a carbon source. Luria-Bertani broth (1 liter; 10 g NaCl, 5 g yeast extract, 10 g tryptone) was also used.

Standard growth conditions. *S. solfataricus* P2 was grown at 80°C (pH 3.5) in Brock's salt medium supplemented with yeast extract, sucrose, and Casamino Acids (0.1% each) (62). The growth of cells was monitored spectrophotometrically at 600 nm, and the cells were harvested at the early stationary phase (1.0 optical density) by centrifugation at 5,000 $\times$ *g* for 20 min at 4°C.

Native protein purification. A culture of 9.0 liters of *S. solfataricus* P2 was centrifuged, and the cellular pellet (9.5 g) was resuspended in 2 ml g^{-1} cells of 50 mM Tris-HCl buffer (pH 8.0) supplemented with 0.1% Triton X-100. The cells were incubated with lysozyme and Benzonase for 60 min at 37°C, lysed mechanically with five cycles of French press, and centrifuged at 10,000 $\times$ *g* for 30 min at 4°C. The FCE was loaded onto a HiLoad 16/10 Q-Sepharose high-performance column (GE Healthcare) equilibrated in 50 mM Tris-HCl buffer (pH 8.0; buffer A) at a flow rate of 2 ml min^{-1}. The run was performed with an initial extensive wash with buffer A (two column volumes), followed by a linear ionic strength gradient from 0 to 1 M NaCl in buffer A (three column volumes) and a final treatment with 1 M NaCl in buffer A (two

column volumes). Under these conditions, the de-*N*-acetylase activity was found to be primarily in the fractions eluted at about 300 mM NaCl. Active fractions were pooled and dialyzed versus 20 mM phosphate buffer (pH 7.0). The dialyzed pool was equilibrated in 1 M ammonium sulfate and loaded onto a HiLoad 26/10 Phenyl Sepharose high-performance column (GE Healthcare), equilibrated at a flow rate of 3 ml min^{-1} with phosphate buffer (pH 7.0) supplemented with 1 M ammonium sulfate. After one column volume of loading buffer, the protein was eluted with a two-step gradient of phosphate buffer (pH 7.0; 0 to 80%, two column volumes; 80 to 100%, three column volumes), followed by a final step at 100% of phosphate buffer (pH 7.0; two column volumes). The protein was eluted in about 90% phosphate buffer (pH 7.0). Active fractions were pooled, dialyzed against PBS buffer (20 mM sodium phosphate buffer [pH 7.3], 150 mM NaCl), and concentrated by ultrafiltration with an Ultracon 10K (cutoff, 10,000 Da). After concentration, the sample was loaded on a Superdex 75HR 10/300 gel filtration column (GE Healthcare), and the run was performed at a flow rate of 0.5 ml min^{-1} in PBS buffer. Active fractions were pooled and concentrated. The de-*N*-acetylase activity was monitored by performing an enzymatic assay on the chromogenic substrate 4Np-α-GlcNAc at 65°C for 16 h. The enzymatic activity was detected presumably due to the absorbance of the secondary product 2-[*N*-(2-nitrophenyl)amino]-2-deoxy-D-glucopyranose [α-GlcNH(2Np)] at 420 nm. After this procedure, SSO2901 was determined to be >95% pure by SDS-PAGE stained with SYPRO Orange.

Enzymatic assay of native SSO2901. The native SSO2901 was assayed in 50 mM phosphate buffer at pH 6.5 for 16 h at 65°C on the different substrates, 4Np-α-GlcNAc and GlcNAc, at the concentration of 5 mM in a final volume of 0.2 ml. The reactions were blocked in ice by adding 0.8 ml of 1 M sodium carbonate (pH 10.2). The absorbance was measured at 420 nm at room temperature. Spontaneous hydrolysis of the substrates was subtracted by using blank mixtures without enzyme, which were prepared and incubated as described above. Analysis of the reaction products was performed using HPAEC-PAD with a PA200 column (Dionex). HPAEC-PAD analyses were performed at a flow rate of 0.5 ml min^{-1} using isocratic elution of 10 mM NaOH for 20 min.

Cloning of *sso2900* and *sso2901*. The *sso2900* and *sso2901* genes were amplified by PCR from the genome of *S. solfataricus* P2 (180 ng) using the following synthetic oligonucleotides: TOPOFWD2900 (5′-CACCATGGATATTTTAGCAGTAGTT-3′), TOPOREV2900His (5′-CTTTTTTGCAAAGGCTTTTATATCTT-3′), TOPOFWD2901 (5′-CACCATGCAAAAAAGTAGTTGTAT-3′), and TOPOREV2901His (5′-ACCTTTAGTCCAAAA AACCTCGT-3′) for *sso2900* and *sso2901*, respectively. The amplification reaction was performed with PfuUltra II Fusion HS DNA polymerase (Stratagene) according to the following program: a hot start of 5 min at 95°C; followed by 10 cycles of 1 min at 95°C, 1 min at 50°C, and 1 min at 72°C; followed by 30 cycles of 1 min at 95°C, 1 min at 55°C, and 1 min at 72°C, followed finally by extension for 10 min at 72°C.

The PCR products were purified by PCR Kleen Spin columns (Bio-Rad) and ligated into expression vector pET101/d-TOPO according to the instructions of the manufacturer (Invitrogen). The ORFs are under the control of the IPTG (isopropyl-β-D-thiogalactopyranoside)-inducible T7 RNA polymerase promoter, and the C-terminal of the proteins were fused to V5 epitope and a 6×His tag. Positive clones were selected through the PCR colony and verified by sequencing.

Heterologous expression and purification of rSSO2900 and rSSO2901. *E. coli* BL21 Star(DE3) cells were grown at 37°C in 0.5 liter of Luria-Bertani broth supplemented with ampicillin (50 μg ml^{-1}). The expression of rSSO2900 was induced by adding 0.1 mM IPTG when growth reached an optical density at 600 nm (OD_{600}) of 1, while rSSO2901 was induced by adding 0.5 mM IPTG when growth reached an OD_{600} of 0.6.

Cell growths were allowed to proceed for 16 and 2 h, for rSSO2900 and rSSO2901, respectively. The cells were harvested by centrifugation at 5,000 × *g*. The resulting cellular pellets were resuspended in 50 mM sodium phosphate buffer (pH 8.0)–300 mM NaCl with a ratio of 5 ml g^{-1} cells, followed by incubation at 37°C for 1 h with 20 mg of lysozyme (Fluka) and Benzonase (Novagen) at 25 U g^{-1} cell. The cells were lysed by French press treatment, and cell debris was removed by centrifugation at 10,000 × *g* for 30 min. The free cell extracts were loaded onto Protino Ni-IDA packed columns (Macherey-Nagel) as recommended by the manufacturer. The eluted fractions were pooled and dialyzed against 1× PBS (pH 7.3) for rSSO2900 and against 25 mM HEPES-NaOH (pH 7.5) for rSSO2901.

Molecular mass determination of rSSO2901. The molecular mass of rSSO2901 was determined by gel filtration on a Superdex 75HR 10/300 FPLC column (GE Healthcare). The molecular weight markers were bovine serum albumin (66 kDa), RNase A (13.7 kDa), aprotinin (6.5 kDa), and vitamin B_{12} (1.3 kDa).

NMR spectroscopy. ^{1}H and ^{13}C NMR spectra were recorded using a Bruker Avance 600 MHz spectrometer equipped with a cryoprobe. All two-dimensional homo- and heteronuclear experiments (double-quantum-filtered correlation spectroscopy [DQF-COSY]; total correlation spectroscopy [TOCSY]; rotating-frame Overhauser spectroscopy [ROESY]; distortionless enhancement by polarization transfer-heteronuclear single quantum coherence [^{1}H-^{13}C DEPT-HSQC]) were performed using the standard pulse sequences available in the Bruker software. Chemical shifts were measured at 298 K in D_2O.

Fluorescamine-based de-*N*-acetylase activity assay. Purified rSSO2901 was tested for de-*N*-acetylase activity using fluorescamine fluorophore. An assay mixture containing 50 mM HEPES (pH 8.5) and 20 mM GlcNAc was preequilibrated at 70°C prior to the addition of rSSO2901 (3 μg). After 5 min, the reaction mixture (150 μl) was quenched by the addition of 20% trichloroacetic acid (50 μl), which precipitates the protein. After centrifugation (16,000 × *g*, 10 min), 125 μl of supernatant was diluted with 375 μl of 1 M borate (pH 9.0), and reacted with 10 mM fluorescamine, (150 μl in CH_3CN; Invitrogen). After 10 min, the resulting fluorescence was measured (excitation, 395 nm; emission, 495 nm) using an FP-8600 fluorescence spectrometer (Jasco). The observed increase in fluorescence (fluorescence units min^{-1}) was converted into μmol/min using a glucosamine (GlcN) standard curve (50 to 150 μM) (63). One enzymatic unit is defined as the amount of enzyme catalyzing the conversion of 1 μmol of substrate into product

in 1 min under standard conditions. To determine the steady-state parameters, the activity was measured at 18 different concentrations of GlcNAc (0 to 80 mM), and the parameters k_{cat}, K_m, and k_{cat}/K_m were obtained by fitting the Michaelis-Menten equation to the initial linear velocities using the curve-fitting program GraphPad Prism.

Temperature, pH, and metal effects. The temperature profile of rSSO2901 activity was determined at 40 to 85°C on 20 mM GlcNAc in 50 mM HEPES buffer (pH 8.5) for 5 min. The thermal stability was evaluated by incubating pure rSSO2901 in HEPES buffer (pH 7.5) at different temperatures (60, 70, and 80°C). At various time intervals, 3-μl aliquots (3 μg of enzyme) were recovered, transferred on ice, and assayed under the conditions described above. The residual activity is expressed as a percentage of the maximal enzymatic activity measured before incubation at the indicated temperatures; the pH optimum was determined by assaying rSSO2901 in 50 mM concentrations of different buffers (citrate/phosphate [pH 4.0 to 6.0]; phosphate [pH 6.0 to 7.5]; HEPES [pH 7.5-8.5]; borate [pH 7.5 to 9.0]; CHES [pH 8.5 to 10]) on 20 mM GlcNAc at 70°C for 5 min. The effect of metal and EDTA on the activity of rSSO2901 was analyzed for Zn^{2+}, Mg^{2+}, Mn^{2+}, Co^{2+}, and EDTA. The active enzyme was assayed under standard conditions in the presence of a 1 mM concentration of each metal ion or EDTA.

Substrate specificity. Enzymatic assays of rSSO2901 on different substrates (GlcNAc, 20 mM; GalNAc, 20 mM; ManNAc, 20 mM; GlcNAc-1P, 5 mM; GlcNAc-6S, 5 mM; UDP-GlcNAc, 20 mM) were performed by using 3 μg of enzyme at 70°C in 50 mM HEPES buffer (pH 8.5) for 5 min for GlcNAc and 16 h for the other substrates. Spontaneous hydrolysis of the substrates was subtracted by using appropriate blank mixtures without enzyme. The reaction products were analyzed by fluorescamine assay and HPAEC-PAD, as described above, and by TLC. TLC was performed on silica gel 60 F254 using *n*-butanol/methanol/ammonia/water (5:4:2:1 [vol/vol]) and *n*-butanol/MetOH/ammonium hydroxide/H_2O (5:4:2:1) as the eluant. The spots were visualized by staining with 1% ninhydrin, followed by incubation at 150°C for 5 min. rSSO2900 was assayed on 10 mM UDP-GlcNAc in 50 mM Tris-HCl (pH 7.5) with 30 μg of enzyme for 16 h at 65°C. HPAEC-PAD analyses were performed at a flow rate of 0.5 ml min^{-1} in an isocratic elution of 10 mM NaOH for 5 min, followed by two further isocratic steps (15 mM NaOH for 5 min and 70 mM NaOH for 40 min).

Steady-state kinetic constants on MU-α-GlcNAc. The enzymatic assays of rSSO2901 on MU-α-GlcNAc were performed by taking advantage of the cooperative action of the de-*N*-acetylase and MalA on this substrate. A 10-μg portion of rSSO2901 was incubated on increasing concentrations of MU-α-GlcNAc (0.1 to 5 mM) under standard conditions. The reactions were stopped in ice after 5 min. An aliquot of rSSO2901 reaction (50 μl) was buffered in 200 mM sodium acetate (pH 5.5), followed by incubation with 10 μg of MalA at 70°C for 1 h. Under these conditions, MalA performed a complete hydrolysis of MU-α-GlcN, produced by SSO2901, in GlcN. The assays were performed in duplicate. Spontaneous hydrolysis of the substrate was subtracted by using appropriate blank mixtures without enzymes, and appropriate control mixtures were performed by incubating the substrates with each enzyme separately. The glucosamine produced by activity on MU-α-GlcNAc was monitored and quantified by measuring the fluorescence of methylumbelliferone at λ_{ex} 360 nm and λ_{em} 449 nm.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.01879-18.

SUPPLEMENTAL FILE 1, PDF file, 4.2 MB.

ACKNOWLEDGMENTS

We are grateful to Emiliano Bedini for the interpretation of NMR spectra and to Mauro Di Fenza for the critical reading of the manuscript.

This study was supported by the ASI project ECMB 2014-026-R.0.

REFERENCES

1. Nicolaus B, Kambourova M, Oner ET. 2010. Exopolysaccharides from extremophiles: from fundamentals to biotechnology. Environ Technol 31:1145–1158. https://doi.org/10.1080/09593330903552094.
2. Eichler J, Koomey M. 2017. Sweet new roles for protein glycosylation in prokaryotes. Trends Microbiol 25:662–672. https://doi.org/10.1016/j.tim.2017.03.001.
3. Koerdt A, Gödeke J, Berger J, Thormann KM, Albers S-V. 2010. Crenarchaeal biofilm formation under extreme conditions. PLoS One 5:e14104. https://doi.org/10.1371/journal.pone.0014104.
4. Albers S, Eichler J, Aebi M. 2015. *Archaea*, p 283–292. *In* Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, Seeberger PH (ed), Essentials of glycobiology. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. https://doi.org/10.1101/glycobiology.3e.022.
5. Jarrell KF, Jones GM, Nair DB. 2010. Biosynthesis and role of N-linked glycosylation in cell surface structures of archaea with a focus on flagella and S layers. Int J Microbiol 2010:470138. https://doi.org/10.1155/2010/470138.
6. Calo D, Eilam Y, Lichtenstein RG, Eichler J. 2010. Towards glycoengineering in archaea: replacement of *Haloferax volcanii* AglD with homologous glycosyltransferases from other halophilic *Archaea*. Appl Environ Microbiol 76:5684–5692. https://doi.org/10.1128/AEM.00681-10.
7. Guan Z, Naparstek S, Calo D, Eichler J. 2012. Protein glycosylation as an adaptive response in Archaea: growth at different salt concentrations leads to alterations in *Haloferax volcanii* S-layer glycoprotein N-glycosylation. Environ Microbiol 14:743–753. https://doi.org/10.1111/j.1462-2920.2011.02625.x.
8. Yurist-Doutsch S, Magidovich H, Ventura VV, Hitchen PG, Dell A, Eichler J. 2010. N-glycosylation in Archaea: on the coordinated actions of *Haloferax volcanii* AglF and AglM. Mol Microbiol 75:1047–1058. https://doi.org/10.1111/j.1365-2958.2009.07045.x.
9. Tripepi M, You J, Temel S, Onder O, Brisson D, Pohlschroder M. 2012.

N-glycosylation of *Haloferax volcanii* flagellins requires known Agl proteins and is essential for biosynthesis of stable flagella. J Bacteriol 194:4876–4887. https://doi.org/10.1128/JB.00731-12.
10. Jarrell KF, Ding Y, Meyer BH, Albers SV, Kaminski L, Eichler J. 2014. N-linked glycosylation in Archaea: a structural, functional, and genetic analysis. Microbiol Mol Biol Rev 78:304–341. https://doi.org/10.1128/MMBR.00052-13.
11. Paul G, Wieland F. 1987. Sequence of the halobacterial glycosaminoglycan. J Biol Chem 262:9587–9593.
12. Frols S. 2013. Archaeal biofilms: widespread and complex. Biochem Soc Trans 41:393–398. https://doi.org/10.1042/BST20120304.
13. Larkin A, Imperiali B. 2011. The expanding horizons of asparagine-linked glycosylation. Biochemistry 50:4411–4426. https://doi.org/10.1021/bi200346n.
14. Eichler J. 2013. Extreme sweetness: protein glycosylation in archaea. Nat Rev Microbiol 11:151–156. https://doi.org/10.1038/nrmicro2957.
15. Kobayashi T, Nishizaki R, Ikezawa H. 1997. The presence of GPI-linked protein(s) in an archaeobacterium, *Sulfolobus acidocaldarius*, closely related to eukaryotes. Biochim Biophys Acta 1334:1–4. https://doi.org/10.1016/S0304-4165(96)00099-2.
16. Nicolaus B, Manca MC, Ramano I, Lama L. 1993. Production of an exopolysaccharide from two thermophilic archaea belonging to the genus *Sulfolobus*. FEMS Microbiol Lett 109:203–206. https://doi.org/10.1111/j.1574-6968.1993.tb06168.x.
17. Zahringer U, Moll H, Hettmann T, Knirel YA, Schafer G. 2000. Cytochrome $b_{558/566}$ from the archaeon *Sulfolobus acidocaldarius* has a unique Asn-linked highly branched hexasaccharide chain containing 6-sulfoquinovose. Eur J Biochem 267:4144–4149. https://doi.org/10.1046/j.1432-1327.2000.01446.x.
18. Palmieri G, Balestrieri M, Peter-Katalinić J, Pohlentz G, Rossi M, Fiume I, Pocsfalvi G. 2013. Surface-exposed glycoproteins of hyperthermophilic *Sulfolobus solfataricus* P2 show a common N-glycosylation profile. J Proteome Res 12:2779–2790. https://doi.org/10.1021/pr400123z.
19. She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CC, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PM, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, Van der Oost J. 2001. The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. Proc Natl Acad Sci U S A 98:7835–7840. https://doi.org/10.1073/pnas.141222098.
20. Ferrara MC, Cobucci-Ponzano B, Carpentieri A, Henrissat B, Rossi M, Amoresano A, Moracci M. 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. https://doi.org/10.1016/j.bbagen.2013.09.022.
21. Cobucci-Ponzano B, Conte F, Strazzulli A, Capasso C, Fiume I, Pocsfalvi G, Rossi M, Moracci M. 2010. The molecular characterization of a novel GH38 alpha-mannosidase from the crenarchaeon *Sulfolobus solfataricus* revealed its ability in de-mannosylating glycoproteins. Biochimie 92:1895–1907. https://doi.org/10.1016/j.biochi.2010.07.016.
22. Cobucci-Ponzano B, Aurilia V, Riccio G, Henrissat B, Coutinho PM, Strazzulli A, Padula A, Corsaro MM, Pieretti G, Pocsfalvi G, Fiume I, Cannio R, Rossi M, Moracci M. 2010. A new archaeal β-glycosidase from *Sulfolobus solfataricus*. J Biol Chem 285:20691–20703. https://doi.org/10.1074/jbc.M109.086470.
23. Cobucci-Ponzano B, Conte F, Rossi M, Moracci M. 2008. The α-L-fucosidase from *Sulfolobus solfataricus*. Extremophiles 12:61–68. https://doi.org/10.1007/s00792-007-0105-y.
24. Sunna A, Moracci M, Rossi M, Antranikian G. 1997. Glycosyl hydrolases from hyperthermophiles. Extremophiles 1:2–13. https://doi.org/10.1007/s007920050009.
25. Moracci M, Ciaramella M, Nucci R, Pearl LH, Sanderson I, Trincone A, Rossi M. 1994. Thermostable β-glycosidase from *Sulfolobus solfataricus*. Biocatalysis 11:89–103. https://doi.org/10.3109/10242429409034380.
26. Peyfoon E, Meyer B, Hitchen PG, Panico M, Morris HR, Haslam SM, Albers SV, Dell A. 2010. The S-layer glycoprotein of the crenarchaeote *Sulfolobus acidocaldarius* is glycosylated at multiple sites with chitobiose-linked N-glycans. Archaea 2010. https://doi.org/10.1155/2010/754101.
27. Albers SV, Meyer BH. 2011. The archaeal cell envelope. Nat Rev Microbiol 9:414–426. https://doi.org/10.1038/nrmicro2576.
28. Johnson JW, Fisher JF, Mobashery S. 2013. Bacterial cell-wall recycling. Ann N Y Acad Sci 1277:54–75. https://doi.org/10.1111/j.1749-6632.2012.06813.x.
29. Koerdt A, Jachlewski S, Ghosh A, Wingender J, Siebers B, Albers SV. 2012. Complementation of *Sulfolobus solfataricus* PBL2025 with an alpha-mannosidase: effects on surface attachment and biofilm formation. Extremophiles 16:115–125. https://doi.org/10.1007/s00792-011-0411-2.
30. Newton GL, Av-Gay Y, Fahey RC. 2000. *N*-Acetyl-1-D-myo-inosityl-2-amino-2-deoxy-alpha-D-glucopyranoside deacetylase (MshB) is a key enzyme in mycothiol biosynthesis. J Bacteriol 182:6958–6963. https://doi.org/10.1128/JB.182.24.6958-6963.2000.
31. Tanaka T, Fukui T, Fujiwara S, Atomi H, Imanaka T. 2004. Concerted action of diacetylchitobiose deacetylase and exo-β-D-glucosaminidase in a novel chitinolytic pathway in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. J Biol Chem 279:30021–30027. https://doi.org/10.1074/jbc.M314187200.
32. Zou Y, Brunzelle JS, Nair SK. 2008. Crystal structures of lipoglycopeptide antibiotic deacetylases: implications for the biosynthesis of A40926 and teicoplanin. Chem Biol 15:533–545. https://doi.org/10.1016/j.chembiol.2008.05.009.
33. Otsuka Y, Yamamoto T, Fukase K. 2017. Syntheses of N-aryl-protected glucosamines and their stereoselectivity in chemical glycosylations. Tetrahedron Lett 58:3019–3023. https://doi.org/10.1016/j.tetlet.2017.06.037.
34. Bem M, Badea F, Draghici C, Caproiu MT, Vasilescu M, Voicescu M, Pencu G, Beteringhe A, Maganu M, Covaci IC, Constantinescu T, Balaban AT. 2008. 4-(D-Glucosamino)-7-nitrobenzoxadiazole: synthesis, anomers, spectra, TLC behavior, and applications. ARKIVOC 2008:218–234. https://doi.org/10.3998/ark.5550190.0009.224.
35. Wang Y, Song JZ, Yang Q, Liu ZH, Huang XM, Chen Y. 2010. Cloning of a heat-stable chitin deacetylase gene from *Aspergillus nidulans* and its functional expression in *Escherichia coli*. Appl Biochem Biotechnol 162:843–854. https://doi.org/10.1007/s12010-009-8772-z.
36. Rolfsmeier M, Haseltine C, Bini E, Clark A, Blum P. 1998. Molecular characterization of the α-glucosidase gene (*malA*) from the hyperthermophilic archaeon *Sulfolobus solfataricus*. J Bacteriol 180:1287–1295.
37. Moracci M, Cobucci Ponzano B, Trincone A, Fusco S, De Rosa M, van Der Oost J, Sensen CW, Charlebois RL, Rossi M. 2000. Identification and molecular characterization of the first alpha-xylosidase from an archaeon. J Biol Chem 275:22082–22089. https://doi.org/10.1074/jbc.M910392199.
38. Gomez H, Lluch JM, Masgrau L. 2013. Substrate-assisted and nucleophilically assisted catalysis in bovine α1,3-galactosyltransferase: mechanistic implications for retaining glycosyltransferases. J Am Chem Soc 135:7053–7063. https://doi.org/10.1021/ja4024447.
39. Smith TK, Crossman A, Borissow CN, Paterson MJ, Dix A, Brimacombe JS, Ferguson MA. 2001. Specificity of GlcNAc-PI de-*N*-acetylase of GPI biosynthesis and synthesis of parasite-specific suicide substrate inhibitors. EMBO J 20:3322–3332. https://doi.org/10.1093/emboj/20.13.3322.
40. Viars S, Valentine J, Hernick M. 2014. Structure and function of the LmbE-like superfamily. Biomolecules 4:527–545. https://doi.org/10.3390/biom4020527.
41. Broadley SG, Gumbart JC, Weber BW, Marakalala MJ, Steenkamp DJ, Sewell BT. 2012. A new crystal form of MshB from *Mycobacterium tuberculosis* with glycerol and acetate in the active site suggests the catalytic mechanism. Acta Crystallogr D Biol Crystallogr 68:1450–1459. https://doi.org/10.1107/S090744491203449X.
42. Hernick M, Fierke CA. 2005. Zinc hydrolases: the mechanisms of zinc-dependent deacetylases. Arch Biochem Biophys 433:71–84. https://doi.org/10.1016/j.abb.2004.08.006.
43. Urbaniak MD, Crossman A, Chang T, Smith TK, van Aalten DM, Ferguson MA. 2005. The *N*-acetyl-D-glucosaminylphosphatidylinositol de-*N*-acetylase of glycosylphosphatidylinositol biosynthesis is a zinc metalloenzyme. J Biol Chem 280:22831–22838. https://doi.org/10.1074/jbc.M502402200.
44. Deli A, Koutsioulis D, Fadouloglou VE, Spiliotopoulou P, Balomenou S, Arnaouteli S, Tzanodaskalaki M, Mavromatis K, Kokkinidis M, Bouriotis V. 2010. LmbE proteins from *Bacillus cereus* are de-N-acetylases with broad substrate specificity and are highly similar to proteins in *Bacillus anthracis*. FEBS J 277:2740–2753. https://doi.org/10.1111/j.1742-4658.2010.07691.x.
45. Barb AW, Zhou P. 2008. Mechanism and Inhibition of LpxC: an essential zinc-dependent deacetylase of bacterial lipid A synthesis. Curr Pharm Biotechnol 9:9–15.
46. Nakamura N, Inoue N, Watanabe R, Takahashi M, Takeda J, Stevens VL, Ki-

noshita T. 1997. Expression cloning of PIG-L, a candidate *N*-acetylglucosaminyl-phosphatidylinositol deacetylase. J Biol Chem 272:15834–15840. https://doi.org/10.1074/jbc.272.25.15834.
47. Watanabe R, Ohishi K, Maeda Y, Nakamura N, Kinoshita T. 1999. Mammalian PIG-L and its yeast homologue Gpi12p are *N*-acetylglucosaminyl-phosphatidylinositol de-*N*-acetylases essential in glycosylphosphatidylinositol biosynthesis. Biochem J 339:185–192. https://doi.org/10.1042/bj3390185.
48. Chang T, Milne KG, Guther ML, Smith TK, Ferguson MA. 2002. Cloning of *Trypanosoma brucei* and *Leishmania major* genes encoding the GlcNAc-phosphatidylinositol de-N-acetylase of glycosylphosphatidylinositol biosynthesis that is essential to the African sleeping sickness parasite. J Biol Chem 277:50176–50182. https://doi.org/10.1074/jbc.M208374200.
49. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178.
50. Pedone E, Limauro D, D'Alterio R, Rossi M, Bartolucci S. 2006. Characterization of a multifunctional protein disulfide oxidoreductase from *Sulfolobus solfataricus*. FEBS J 273:5407–5420. https://doi.org/10.1111/j.1742-4658.2006.05533.x.
51. Gaballa A, Newton GL, Antelmann H, Parsonage D, Upton H, Rawat M, Claiborne A, Fahey RC, Helmann JD. 2010. Biosynthesis and functions of bacillithiol, a major low-molecular-weight thiol in *Bacilli*. Proc Natl Acad Sci U S A 107:6482–6486. https://doi.org/10.1073/pnas.1000928107.
52. Newton GL, Buchmeier N, Fahey RC. 2008. Biosynthesis and functions of mycothiol, the unique protective thiol of *Actinobacteria*. Microbiol Mol Biol Rev 72:471–494. https://doi.org/10.1128/MMBR.00008-08.
53. Hernick M. 2013. Mycothiol: a target for potentiation of rifampin and other antibiotics against *Mycobacterium tuberculosis*. Expert Rev Anti Infect Ther 11:49–67. https://doi.org/10.1586/eri.12.152.
54. Hou J, Han J, Cai L, Zhou J, Lu Y, Jin C, Liu J, Xiang H. 2014. Characterization of genes for chitin catabolism in *Haloferax mediterranei*. Appl Microbiol Biotechnol 98:1185–1194. https://doi.org/10.1007/s00253-013-4969-8.
55. Zurzolo C, Simons K. 2016. Glycosylphosphatidylinositol-anchored proteins: membrane organization and transport. Biochim Biophys Acta 1858:632–639. https://doi.org/10.1016/j.bbamem.2015.12.018.
56. Low MG. 1989. The glycosyl-phosphatidylinositol anchor of membrane proteins. Biochim Biophys Acta 988:427–454. https://doi.org/10.1016/0304-4157(89)90014-2.
57. Low MG, Saltiel AR. 1988. Structural and functional roles of glycosyl-phosphatidylinositol in membranes. Science 239:268–275. https://doi.org/10.1126/science.3276003.
58. Urbaniak MD, Ferguson MAJ. 2009. The GlcNAc-PI de-N-acetylase: structure, function, and activity, p 49–64. *In* The enzymes, vol 26. Academic Press, Inc., New York, NY.
59. Whitfield GB, Marmont LS, Howell PL. 2015. Enzymatic modifications of exopolysaccharides enhance bacterial persistence. Front Microbiol 6:471. https://doi.org/10.3389/fmicb.2015.00471.
60. DeAngelis PL, Gunay NS, Toida T, Mao WJ, Linhardt RJ. 2002. Identification of the capsular polysaccharides of type D and F *Pasteurella multocida* as unmodified heparin and chondroitin, respectively. Carbohydr Res 337:1547–1552. https://doi.org/10.1016/S0008-6215(02)00219-7.
61. DeAngelis PL, Liu J, Linhardt RJ. 2013. Chemoenzymatic synthesis of glycosaminoglycans: re-creating, re-modeling and re-designing nature's longest or most complex carbohydrate chains. Glycobiology 23:764–777. https://doi.org/10.1093/glycob/cwt016.
62. Brock TD, Brock KM, Belly RT, Weiss RL. 1972. *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch Mikrobiol 84:54–68. https://doi.org/10.1007/BF00408082.
63. Huang X, Hernick M. 2011. A fluorescence-based assay for measuring *N*-acetyl-1-D-myo-inosityl-2-amino-2-deoxy-α-D-glucopyranoside deacetylase activity. Anal Biochem 414:278–281. https://doi.org/10.1016/j.ab.2011.04.001.

molecules

MDPI

Article

Spatial Metagenomics of Three Geothermal Sites in Pisciarelli Hot Spring Focusing on the Biochemical Resources of the Microbial Consortia

Roberta Iacono [1], Beatrice Cobucci-Ponzano [2], Federica De Lise [2], Nicola Curci [1,2], Luisa Maurelli [2], Marco Moracci [1,2,3,*] and Andrea Strazzulli [1,3,*]

[1] Department of Biology, University of Naples "Federico II", Complesso Universitario Di Monte S. Angelo, Via Cupa Nuova Cinthia 21, 80126 Naples, Italy; roberta.iacono@unina.it (R.I.); nicola.curci@unina.it (N.C.)

[2] Institute of Biosciences and BioResources, National Research Council of Italy, Via P. Castellino 111, 80131 Naples, Italy; beatrice.cobucciponzano@ibbr.cnr.it (B.C.-P.); federica.delise@ibbr.cnr.it (F.D.L.); luisa.maurelli@ibbr.cnr.it (L.M.)

[3] Task Force on Microbiome Studies, University of Naples Federico II, 80134 Naples, Italy

* Correspondence: marco.moracci@unina.it (M.M.); andrea.strazzulli@unina.it (A.S.)

Academic Editor: Raffaele Saladino

Received: 11 August 2020; Accepted: 28 August 2020; Published: 3 September 2020

Abstract: Terrestrial hot springs are of great interest to the general public and to scientists alike due to their unique and extreme conditions. These have been sought out by geochemists, astrobiologists, and microbiologists around the globe who are interested in their chemical properties, which provide a strong selective pressure on local microorganisms. Drivers of microbial community composition in these springs include temperature, pH, in-situ chemistry, and biogeography. Microbes in these communities have evolved strategies to thrive in these conditions by converting hot spring chemicals and organic matter into cellular energy. Following our previous metagenomic analysis of Pisciarelli hot springs (Naples, Italy), we report here the comparative metagenomic study of three novel sites, formed in Pisciarelli as result of recent geothermal activity. This study adds comprehensive information about phylogenetic diversity within Pisciarelli hot springs by peeking into possible mechanisms of adaptation to biogeochemical cycles, and high applicative potential of the entire set of genes involved in the carbohydrate metabolism in this environment (CAZome). This site is an excellent model for the study of biodiversity on Earth and biosignature identification, and for the study of the origin and limits of life.

Keywords: origin of life; microbial community; CAZymes; extremozymes; environmental changes; comparative metagenomics

1. Introduction

Extreme environments such as hot springs are of great interest as a source of novel extremophilic microorganisms, enzymes, and metabolic pathways essential for the microbial survival in extreme conditions [1]. Extremophiles are known to thrive in diverse extreme conditions, such as high or low temperatures, high salinity, acidic and alkaline pH values, and high radiation [2]. They not only can tolerate these conditions but require the latter for survival. Exploring the diversity of extremophiles and understanding their mechanisms of adaptation [3] permit us to expand our notions of the potential habitable environments able to sustain life beyond Earth [4]. Indeed, sites harboring harsh environments identified during solar system exploration may host now, or may have hosted, extremophilic life forms.

Research on extremophiles and their enzymes (extremozymes) has not only reshaped our understanding of the origin and evolution of life [5] and the potential for life on other planetary

bodies [6], but also it has simultaneously led to numerous advances in molecular biology, medicine, and biotechnology [7–10]. In fact, extremozymes represent interesting cases of protein adaptation under conditions where conventional enzymes quickly denature [11–13]. Thus, extremozymes are ideal tools for industrial applications where harsh chemical and physical conditions are encountered. However, the difficulties in cultivating extremophiles severely limit access to this class of biocatalysts, thereby metagenomic approaches are now largely used for extremozyme discovery [14–17].

The increasing amounts of metagenomic data and fully sequenced genomes now allow us to systematically explore these microbial communities [14,18], enabling us to investigate the uncultured microbial population, the mechanisms of possible adaptation to biogeochemical cycles, and the lifestyles of extreme organisms, and to discover new extremozymes [19].

Recently, our group reported on the metagenomic analysis of the microbial community populating the Pisciarelli hot springs (Naples, Italy), identifying the entire carbohydrate active enzymes portfolio (CAZome) that has been cloned and partially characterized [20].

In detail, two main mud/water pools have been identified in Pisciarelli in March 2012, named Pool 1 (T = 85 °C, pH 5.5) and Pool 2 (T = 92 °C, pH 1.5). The first pool was almost exclusively populated by Archaea (*Acidianus hospitalis*, 40%; *Pyrobaculum arsenaticum*, 20%; *Pyrobaculum oguniense*, 5%; *Saccharolobus solfataricus*, 1%) followed by Bacteria (0.11%) and archaeal viruses (0.17%), while more than 30% of the obtained reads found no match with the nucleotide NCBI database. In contrast, the majority of the obtained reads from Pool 2 found no match with the NT NCBI database (62%). Among the assigned reads, Crenarchaeota (37%) dominate the site (*Metallosphaera sedula*, 31%; *Saccharolobus solfataricus*, 3%; *Acidianus hospitalis*, 3%) followed by archaeal viruses (0.36%) [20]. In that study, we demonstrated that even sites that have been consistently sampled for decades are still largely unexplored in terms of microbial diversity and of their extremozymes. The microbial population in Pisciarelli has been shown to have a huge number of genes encoding putative CAZymes, which include glycoside hydrolases (GHs), carbohydrate esterases (CEs), polysaccharide lyases (PLs), and auxiliary activities (AAs) [20]. These activities are classified in the CAZy database (www.cazy.org) [21]. Thus, these biocatalysts are ideal candidates for biotechnological applications and to understand enzyme adaptations to extreme environments.

In the last 15 years, the geothermal activity in the Pisciarelli area has been increasing, showing a rise in fumarolic discharge, the formation of boiling pools and water springs (March 2009), and the opening of energetic geyser-type vents (November 2010) that are currently very active. The temperature at the main Pisciarelli fumarole rose up to 110 °C in 2011, but dropped sharply in April 2012 to the present value of 95 °C [22,23]. In December 2012, the Italian Civil Protection Department raised the alert of the caldera from green level (base) to the current yellow level (attention) as a consequence of a further increase in the deformation rate, seismicity, and degassing [22]. Currently, the hydrothermal activity at Pisciarelli shows an escalation characterized by an increase in the CO_2 flux, which in 2019 exceeded 500 t/day [24].

For these reasons, we embarked on a novel metagenomic study of the Pisciarelli hot springs in order to explore the microbial communities populating mud/water pools formed as a result of the local changes that occurred.

To date, Pisciarelli represents a unique ecological niche for comparative metagenomics studies. The access to this site provides valuable insight into the adaptive strategies of the extremophiles communities, and conditions generally difficult to study in other remote extreme environments and/or to reproduce in lab. The study can lead to a more comprehensive understanding of the mechanisms of evolutionary change that underlie the adaptation of microbes to extreme conditions.

2. Results and Discussion

2.1. Sampling in the Solfatara Hot Springs, mDNA Extraction, and Sequencing

The aim of this study is to explore the microbial communities populating three sites of the Pisciarelli hot springs (40°49′45.1″ N; 14°8′49.4″ E), named Site A, Site B, and Site C (Figure 1A) and investigate their differences in terms biodiversity and potential source of enzymes.

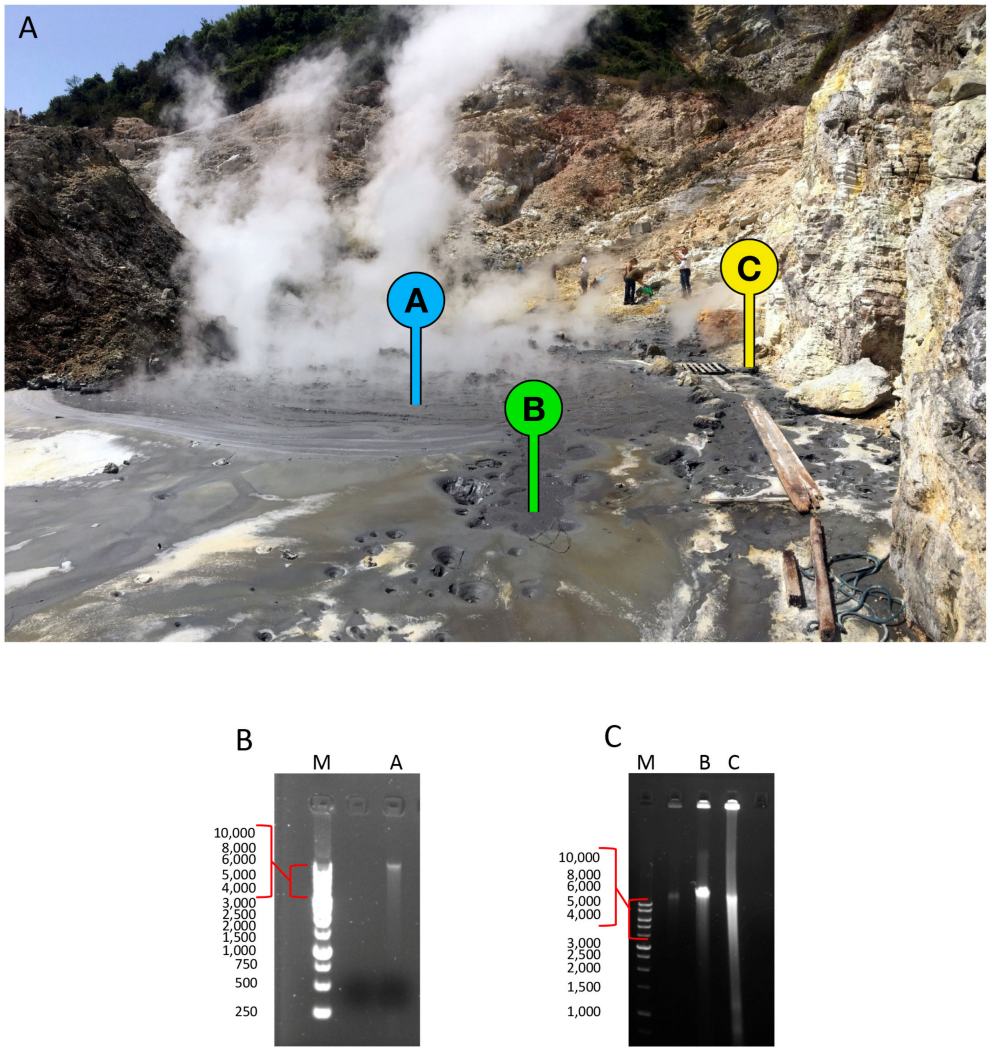


Figure 1. (**A**) View of the sampling site in July 2019. The sites A, B, and C are labeled by badges. (**B**) Agarose gel mDNA extraction from Site A. Lane M, Marker StoS 1Kb DNA Ladder (Genespin), lane A: mDNA from Site A. (**C**) Agarose gel of mDNA extractions from sites B and C. Lane M, Marker StoS 1Kb DNA Ladder (Genespin); lane B, mDNA from Site B; lane C, mDNA from Site C.

Site A (94.1 °C; pH 5.2) is the largest pool of the area, mainly made by water and mud. Recently, the microbial community populating Site A (then called Pool 1) has been characterized through a metagenomic approach showing the dominance of Archaea belonging to the genera *Acidianus* (40%) and *Pyrobaculum* (25%), as well as the presence of sequences related to new phyla ascribable to the

Sulfolobaceae family. A dramatic morphological change in the area of interest, which occurred at the end of July 2019, led to the expansion of the area previously known as Pool 1, which merged to the proximal Pool 2 pool (92 °C, pH 1.5), generating a completely new site (Site A). This new pool, if compared to Pool 1, changed both in terms of extension of the pool boundaries and in terms of temperature, showing, in particular, an increase of about 9 °C compared to the conditions previously observed [20].

Site B (47.7 °C; pH 5.8) is mainly composed of mud and is physically adjacent to Site A from which it receives part of the liquid fraction. Nonetheless, Site B has its emission of gas, which can be observed through the formation of bubbles that contribute to the mixing of the liquid on the surface.

Site C (73.0 °C; pH 2.5) differently, is physically distant from Site A and Site B and is a shallow pool of water located near to the rocky wall of the area (Figure 1A) and characterized by intense steam jets.

Temperature and pH were monitored in-situ at the three sites and the samples, composed of water and sediment, were collected and taken to the laboratory where, by centrifugation, sediments of 17 g, 50 g, and 40 g were obtained from Sites A, B, and C, respectively.

The sediment obtained from each sample was treated for the extraction of the whole mDNA obtaining 27, 60, and 164 ng/g of sediment from Site A, Site B, and Site C, respectively (Figure 1B,C).

The mDNA was then sequenced in outsourcing by Novogene-Europe (Cambridge, UK) through Illumina MiSeq (150 PE), obtaining 23,830,104 clean reads from Site A, 22,933,864 from Site B, and 23,961,446 from Site C.

2.2. Microbial Communities

To evaluate the composition of the microbial communities populating the three sites, the obtained reads were analyzed by *blastn* against the NCBI NT nucleotide database.

The analysis revealed that all three sites are dominated by Archaea, and Site A had the highest number of reads (84%) assigned to this kingdom, followed by Site B (67%) and Site C (53%) (Figure 2). It is worth noting that each site showed a high number of reads that had no match in the NT database (unassigned). Notably, unassigned reads represented 15, 27, and 41% of the whole reads of Site A, B, and C, respectively. Furthermore, Site B had the highest number of reads assigned to the kingdom of Bacteria (5%) compared to the Sites A and C, where less than 1% of sequences could be assigned to this kingdom. Site C, on the contrary, showed a percentage of viral sequences (4.6%) higher than Sites A and B (0.3 and 0.8%, respectively).

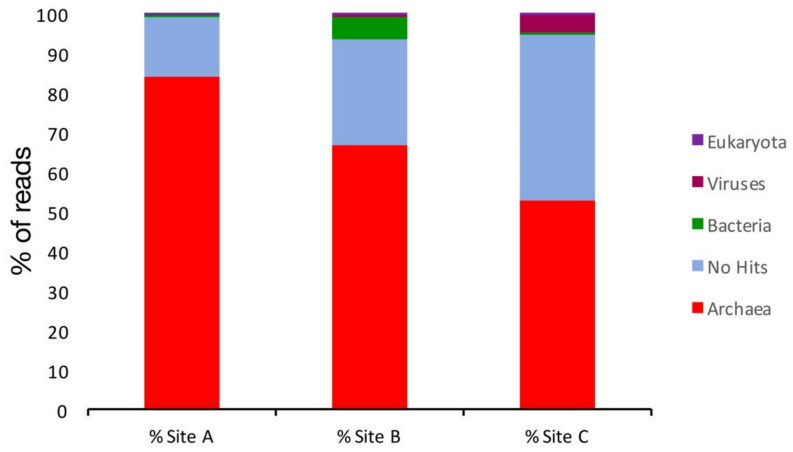


Figure 2. Taxonomic assignment of the reads at the kingdom level.

The detailed investigation of the taxonomically assigned reads, at the level of genus and species (Figures 3 and 4), revealed that Site A is mainly dominated by the genus *Acidianus* (47%), in particular the species *A. ambivalens* (30%) and *A. hospitalis* (16%), followed by the genus *Pyrobaculum* (35%) mainly attributable to the species *P. arsenaticum* (31%). This result differs from what was previously observed on the Pool 1 site in which we observed the dominance of *A. hospitalis* (40%) followed by *P. arsenaticum* (20%) [20].

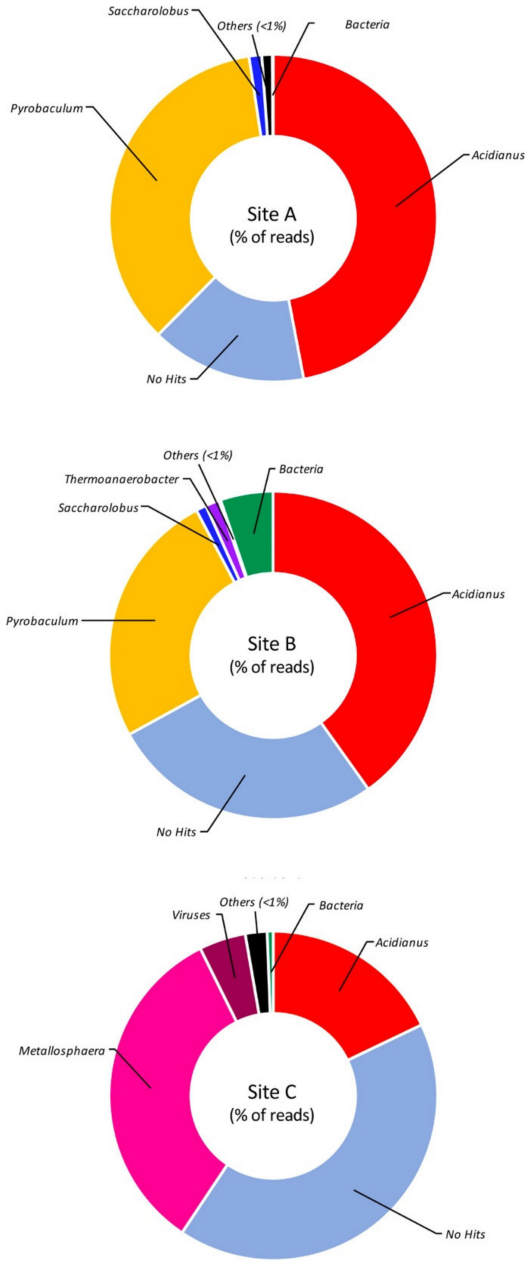


Figure 3. Taxonomic assignment of the reads at the genus level. Taxa showing less than 1% of assigned reads are grouped as "others".

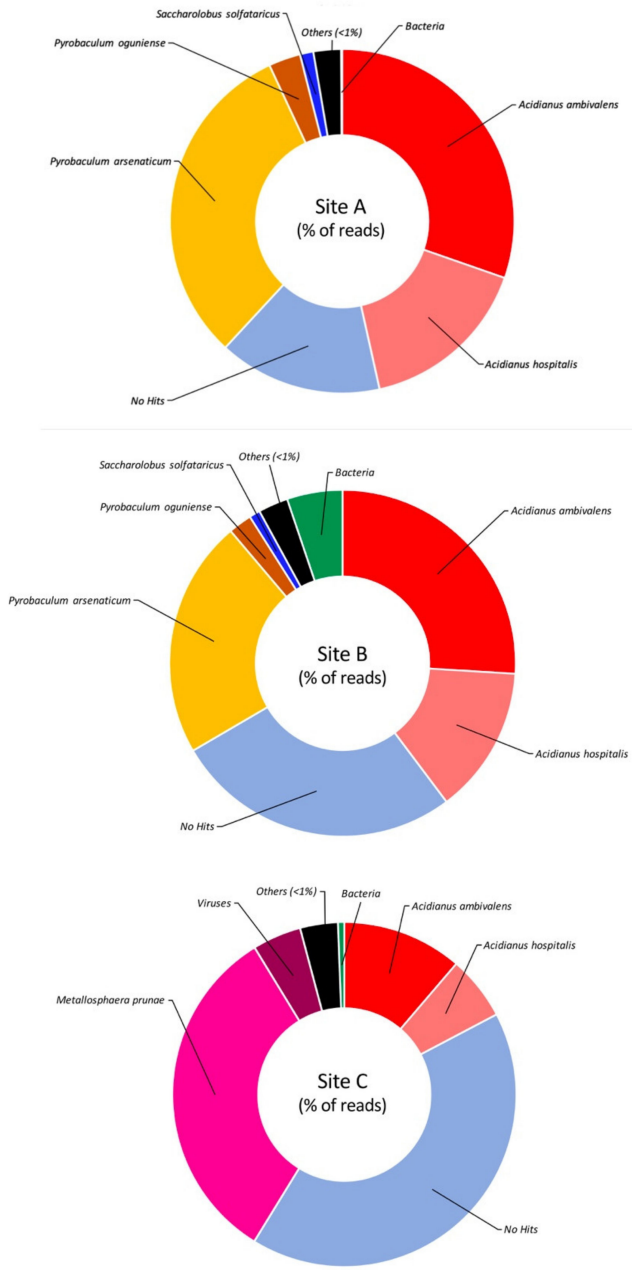


Figure 4. Taxonomic assignment of the reads at the species level. Taxa showing less than 1% of assigned reads are grouped as "others".

A notable difference between the result observed in Site A and the previous study consists in the percentage of reads of unknown origin, which decreased from 32% in Pool 1 to 15.4% in Site A. This variation indicates a change in biodiversity of the microbial community populating the pool as an effect of the geothermal events that occurred between March 2012 and July 2019 (Pool 1 Biosample ID: SAMN09692669).

Site B, alike Site A, also showed *Acidianus* (40%) as the dominant genus, followed by *Pyrobaculum* (25%) (Figure 3) with a relative abundance of species comparable to what was observed in Site A (Figure 4). This parallelism can be explained considering that the two sites are not physically distant to each other and that the liquid fraction present in Site B was partially provided by the proximal Site A (Figure 1A). Indeed, the main difference between Site A and Site B is that in the latter, 5% of the reads were assigned to bacterial origins and the number of sequences not assigned to known microorganisms was 1.7-fold higher than those present in Site A.

Reads assigned to the genus *Acidianus* were also present in the sample from Site C. These, however, represented only 18% of the total reads, while the dominant genus was represented by *Metallosphaera* (33%) (Figure 3), mainly attributable to the species *M. prunae*, which is known to grow at temperatures between 55 and 80 °C and pH range 1.0–4.5 [25]. This result is in line with what was previously observed in another extremely acidic site of the Pisciarelli hot spring, named Pool 2 (92 °C, pH 1.5) (Pool 2 Biosample ID: SAMN09692670), where *Metallosphaera sedula* was the dominant species [20]. Nonetheless, the most abundant component (41.5%) of Site C consisted of reads that did not match any known sequence.

Another remarkable difference between Site C and the other two sites was the number of viral reads identified in this sample. As mentioned above, 4.6% of Site C reads were taxonomically assigned to viruses. Among these, viruses mainly belonging to Bicaudaviridae (different variants of *Acidianus* two-tailed virus and disparate types of *Sulfolobus monocaudavirus*), Ampullaviridae (*Acidianus* bottle-shaped viruses), Fuselloviridae (*Sulfolobus* spindle-shaped viruses), and Ligamenvirales (*Sulfolobales* rod-shaped viruses) were identified.

It is known that viruses play a key role in horizontal gene transfer (HGT) in prokaryotes [26]. Transfer of DNA has been shown to be involved in genome evolution and in adaptation to high temperatures [27]. In particular, it has been proven that spindle-shaped fuselloviruses that infect *Sulfolobus* and *Acidianus* species can promote the virus-mediated HGT between different hosts [28], contributing significantly to the dynamic of the prokaryotic genomes. Thus, the presence in high percentages of viral sequences in Site C might be attributed as a survival mechanism against rapid environmental changing of this extreme site. Therefore, the peculiar microbial composition of Site C is presumably related to the considerably more acidic pH value if compared to the other two mud pools.

2.3. Analysis of Bacteria Communities

To evaluate the bacterial communities present in Sites A, B, and C, the reads of the three samples were analyzed in detail in the NT database (Table 1).

Table 1. Relative abundances of the bacteria communities in the Sites A, B, and C.

Genus	Site A (%)	Site B (%)	Site C (%)	Temperature Range
Thermoanaerobacter	0.10	26.00	1.24	Hyperthermophilic
Thiomonas	0.10	14.00	0.00	Mesophilic/Moderately thermophilic [a]
Thermoanaerobacterium	0.00	10.00	0.00	Thermophilic
Caldanaerobacter	0.00	9.00	1.24	Hyperthermophilic
Acidithiobacillus	15.00	4.00	0.00	Mesophilic/Moderately thermophilic [b]
Clostridium	6.00	0.50	10.00	Mesophilic
Hydrogenobacter	8.00	0.10	0.00	Hyperthermophilic
Aeromonas	0.10	0.10	8.00	Mesophilic
others (< 7% of bacterial reads)	70.70	36.30	79.52	mixed

[a] [29–31]. [b] [32].

In Site A, 22,276 were assigned to bacteria (Figure 5A) whose most abundant genera were represented by the mesophilic/moderately thermophilic bacterium *Acidithiobacillus* (15%) and the

hyperthermophilic *Hydrogenobacter* (8%), while the remaining 78% belonged to different genera whose relative abundance was less than 7% (Figure 5B). Regarding the genus *Acidithiobacillus* it is important to highlight that, although this is mainly represented by mesophilic microorganisms, it also groups the moderately thermophilic *Acidithiobacillus caldus* with an optimal growth pH between 2.0–2.5 and with an optimal temperature of 45 °C [32].

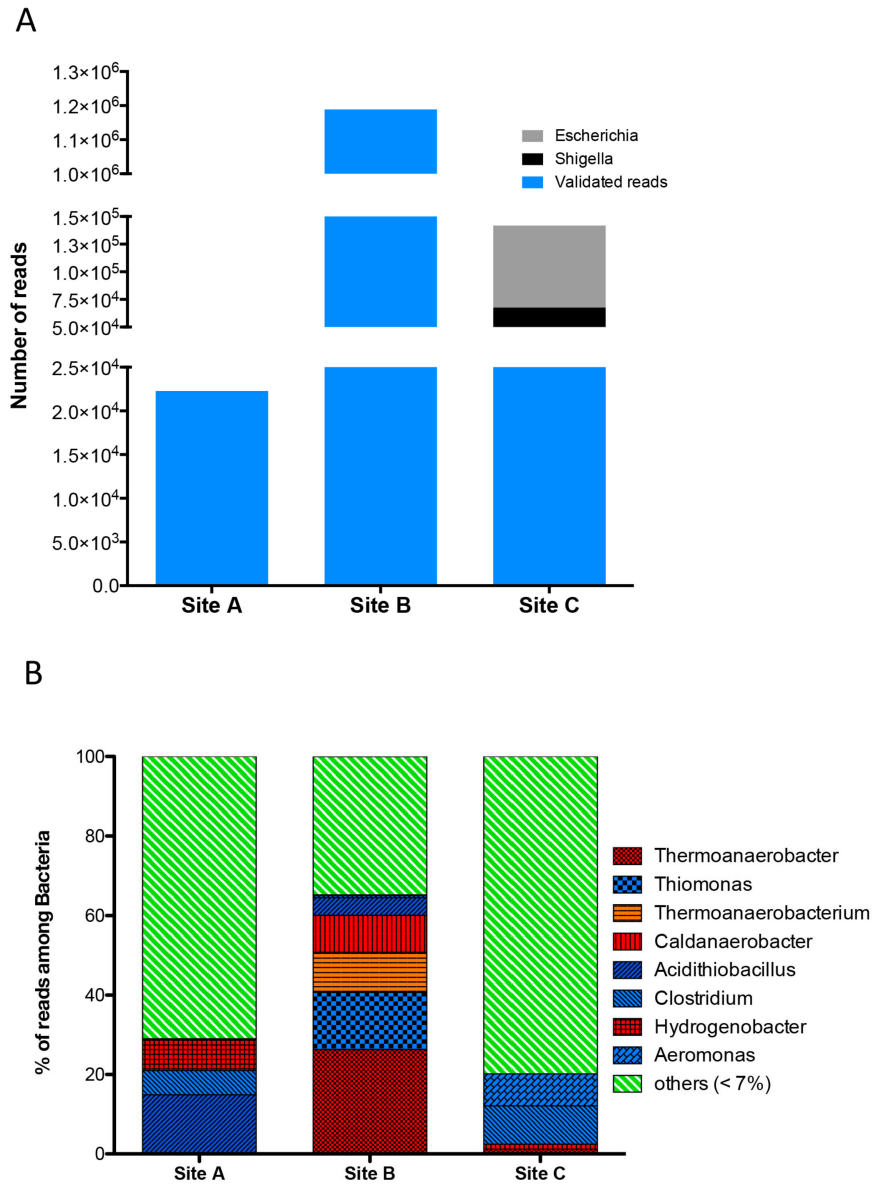


Figure 5. (**A**) Number of reads in Sites A, B, and C assigned to bacteria. Black and grey indicate the reads assigned to the genera *Shigella* and *Escherichia*, respectively, and filtered as contaminants. In blue the number of validated reads used for the bacteria community analysis. (**B**) Bacteria community profile, relative abundances, and diversity: hyperthermophiles (red), thermophiles (orange), mesophiles (blue), others (each relative abundance % < 7, green).

As previously indicated, among the three sites, Site B showed by far the highest percentage of bacterial reads (1,188,674). The analysis of the relative abundances of bacteria present in the microbial community of Site B (Figure 5B) allowed us to identify the hyperthermophiles *Thermoanaerobacter* (26%) and *Caldanaerobacter* (9%), the thermophile *Thermoanaerobacterium* (10%), and *Acidithiobacillus*, which, unlike in Site A, was much less abundant here (4%). It is important to note that one of the most abundant bacterial genus present in Site B was *Thiomonas* (14%), which, although generally grouping mesophilic species, also includes moderately thermophilic species identified in geothermal sources at ~45 °C and able to grow at temperatures up to 50 °C and in the pH range 4.0–7.0 [29–31].

In Site C, 141,784 reads were assigned to the kingdom of Bacteria. Of these, 15 and 52% were assigned to the genera *Shigella* and *Escherichia*, respectively (Figure 5A). Since both are mesophilic gammaproteobacteria whose natural habitat is the human and animal gut [33], these reads were considered as environmental contaminations and not taken into account for the purpose of evaluating the bacterial population.

Among the remaining 46,141 reads of Site C, the most abundant genera were *Clostridium* (10%) and *Aeromonas* (8%). The *Clostridium* genus includes obligate anaerobic bacteria and Gram-positive bacteria and are capable of forming spores in adverse environmental conditions, which populate soil, sand, rivers, swimming pools, river bank mud, and marine sediments [34]. Unfortunately, the low number of reads from Site C, associated with this genus, has not allowed for a more detailed taxonomic annotation, and it is therefore currently impossible to trace the species present in the sample. However, recently, two studies on microbial communities populating Malaysian hot springs (temperatures range 50–110 °C) [35] and five hot springs in Eritrea (temperatures between 45 and 100 °C) [36] revealed the presence of various human pathogens, including *Clostridium spp*. and *Aeromonas*. In addition, a novel thermophilic *Clostridium* species (*C. thermarum*) from a thermal spring in China has recently been identified and characterized [37].

2.4. Assembly, Clustering, and Taxonomic Analysis of Unassigned Reads

To identify possible chunks of individual genomes present in the three samples, all the reads were separately assembled by MEGAHIT [38] obtaining 6296, 38,136, and 16,854 contigs in Sites A, B, and C, respectively (Table 2).

Table 2. Assembly statistics.

	Site A	Site B	Site C
Number of contigs	6296	38,136	16,854
Total contigs length	9,459,744	53,411,732	18,962,457
Mean length	1502	1400	1125
SD value	160	33	54
Max contigs length	485,160	482,809	417,005
Min contigs length	200	200	200
N50 value	24,403	3592	2299
N80 value	737	727	602
N90 value	414	461	382

Contigs with a length ≥1000 bp were analyzed by MyCC [39], thus allowing the identification of 25 clusters in Site A, 21 clusters in Site B, and 16 clusters in Site C (Figure S1).

The clusters obtained were then analyzed by CheckM [40], which allowed us to validate those with completeness values ≥20%, obtaining five clusters in Site A (7, 12, 17, 22, and 23), fifteen clusters in Site B (2, 3, 4, 6, 7, 8, 9, 11, 13, 14, 15, 16, 18, 19, and 20), and five clusters in Site C (1, 4, 5, 8, and 10) (Tables S1 and S2).

To obtain a taxonomic assignment, the validated clusters were analyzed by Diamond (in blastx mode) using the NCBI Refseq Protein database [41] (Table 3). The result of this analysis made it possible to note the Site A clusters as belonging to the *Crenarchaeota* phylum, in particular related to the genera *Acidianus*, *Pyrobaculum*, and *Desulfurococcus*.

Table 3. Taxonomic assignment of Sites A, B, and C CheckM validated clusters.

Sample	Cluster	Phylum	Dominant Taxa
Site A	Cluster 7	Crenarchaeota	*Desulfurococcus*
	Cluster 12	Crenarchaeota	*Pyrobaculum arsenaticum*
	Cluster 17	Crenarchaeota	*Pyrobaculum* spp.
	Cluster 22	Crenarchaeota	*Acidianus hospitalis*
	Cluster 23	Crenarchaeota	*Acidianus* spp.
Site B	Cluster 2	Mixed	mixed
	Cluster 3	Proteobacteria	*Acidithiobacillus caldus*
	Cluster 4	Firmicutes	*Caldanaerobius* spp.
	Cluster 6	Crenarchaeota	*Desulfurococcus* spp.
	Cluster 7	Crenarchaeota	*Sulfolobaceae*
	Cluster 8	Firmicutes	*Thermoanaerobacterium* spp.
	Cluster 9	Firmicutes	*Thermoanaerobacter* spp.
	Cluster 11	Proteobacteria	*Thiomonas* spp.
	Cluster 13	Crenarchaeota	*Pyrobaculum arsenaticum*
	Cluster 14	Firmicutes	*Caldanaerobacter subterraneus*
	Cluster 15	Firmicutes	*Hydrogenibacillus* spp.; *Thermicanus* spp.
	Cluster 16	Firmicutes	*Desulfotomaculum copahuensis*
	Cluster 18	Proteobacteria	*Desulfurella* spp.
	Cluster 19	Firmicutes	*Caldanaerobius* spp.
	Cluster 20	Mixed	mixed
Site C	Cluster 1	Crenarchaeota	*Acidianus brierleyi*
	Cluster 4	Mixed	mixed
	Cluster 5	Mixed	mixed
	Cluster 8	Crenarchaeota	*Acidianus* spp.
	Cluster 10	Crenarchaeota	*Acidianus* spp.

Site C, on the contrary, was characterized by clusters entirely related to *Acidianus spp.* and by two clusters (4 and 5) with a highly heterogeneous assignment that prevented the identification of a dominant species.

Differently, the clusters identified in Site B were mainly assigned to bacteria belonging to the phyla Firmicutes and Proteobacteria, confirming what was already observed from the taxonomic analysis of the reads. Three of the clusters of Site B were assigned to the *Crenarchaeota* phylum, related to *Desulforococcus* spp. and *P. arsenaticum.*

To identify the origin of the reads that had no match against the NCBI NT nucleotide database, these reads were aligned using Bowite2 [42] (Table 4). Regarding the unassigned Site A reads, most of them were aligned with clusters 9, 22, and 23. The first two clusters were assigned to the genus *Acidianus* (Table S1) and presumably represented the result of HGT, as previously reported [20].

Table 4. Percent of reads without match vs. NT aligned to each cluster.

Site A		Site B		Site C	
Cluster	**% of Unassigned Reads Aligned**	**Cluster**	**% of Unassigned Reads Aligned**	**Cluster**	**% of Unassigned Reads Aligned**
Cluster 1	0.59	Cluster 1	0.15	Cluster 1 *	0.51
Cluster 2	2.98	Cluster 2 *	3.71	Cluster 2	0.32
Cluster 3	0.05	Cluster 3 *	0.77	Cluster 3	0.1
Cluster 4	5.02	Cluster 4 *	3.83	Cluster 4 *	0.45
Cluster 5	1.59	Cluster 5	0.36	Cluster 5 *	2.17
Cluster 6	1.07	Cluster 6 *	3.48	Cluster 6	1.58
Cluster 7*	11.9	Cluster 7 *	36.51	Cluster 7	0.48
Cluster 8	0.05	Cluster 8 *	0.43	Cluster 8 *	4.4
Cluster 9	13.93	Cluster 9 *	2.43	Cluster 9	0.07
Cluster 10	1.04	Cluster 10	1.84	Cluster 10 *	85.64
Cluster 11	1.22	Cluster 11 *	8.85	Cluster 11	0.05
Cluster 12 *	4.17	Cluster 12	0.08	Cluster 12	0.68
Cluster 13	8.38	Cluster 13 *	12.92	Cluster 13	0.29
Cluster 14	0.17	Cluster 14 *	0.38	Cluster 14	0.61
Cluster 15	3.71	Cluster 15 *	8.65	Cluster 15	2.43
Cluster 16	0.22	Cluster 16 *	2.62	Cluster 16	0.22
Cluster 17 *	2.79	Cluster 17	2.04		
Cluster 18	5.73	Cluster 18 *	6.75		
Cluster 19	1.23	Cluster 19 *	1.88		
Cluster 20	0.05	Cluster 20 *	2.19		
Cluster 21	0.76	Cluster 21	0.13		
Cluster 22 *	14.89				
Cluster 23 *	18.2				
Cluster 24	0.02				
Cluster 25	0.24				

* Clusters validated by CheckM.

Instead, cluster 23 was assigned to the genus *Hydrogenobacter*, suggesting the possible presence of microorganisms not yet identified belonging to the Aquificaceae family.

As regards the unassigned Site B reads, these were aligned mainly against clusters 7 and 13. While cluster 13 was assigned to *Pyrobaculum arsenaticum,* indicating also in this case a probable HGT event, cluster 7 was instead taxonomically identified only at the family level as Sulfolobaceae.

The unassigned reads of Site C represented a completely special case. Indeed, these mostly aligned (>85%) to cluster 10, whose taxonomic analysis was classified as related to the genus *Acidianus*. Observing the contamination value of cluster 10 (45%) (Table S1), with which CheckM indicated the percentage of the expected number of duplicate single-copy markers, it was legitimate to assume that this cluster had grouped contigs belonging to species of *Acidianus* not yet identified.

2.5. Evaluation of Microbial Replication Rates

To evaluate the individual contribution to the metabolic functions of the microbial consortia present in Sites A, B, and C, the replication indices of the validated clusters were calculated using iRep [43].

Among all the analyzed clusters, only ten respected the selection parameters (Table S3) and were analyzed with iRep, which was able to determine the replication index for only four of these: cluster 7 of Site A and clusters 2, 4, and 6 of Site B (Table 5).

Table 5. iRep indices for validate clusters in sites A, B, and C.

Sample	Cluster	iRep Index	Dominant Taxa
Site A	Cluster 7	1.32	*Desulfurococcus*
Site B	Cluster 11	n/a	*Thiomonas* spp.
	Cluster 13	n/a	*Pyrobaculum arsenaticum*
	Cluster 15	n/a	*Hydrogenibacillus* spp.; *Thermicanus* spp.
	Cluster 18	n/a	*Desulfurella* spp.
	Cluster 2	1.57	Thermoanaerobacteriales
	Cluster 4	1.51	*Caldanaerobius* spp.
	Cluster 6	1.39	*Desulfurococcus* spp.
	Cluster 7	n/a	*Sulfolobaceae*
	Cluster 9	n/a	*Thermoanaerobacter* spp.
Site C	Cluster 5	n/a	mixed
	Cluster 10	n/a	*Acidianus* spp.

n/a: not applicable.

The obtained replication indices allowed us to estimate the percentage of replication of the microbial species associated with the clusters indicating that in Site A and in Site B, *Desulfurococcus*, despite the low number of reads assigned to this taxon (<0.1% of the total reads in both sites), had more than 30% of the cells in active replication (Figure S2).

In addition, regarding Site B, iRep showed that the bacteria belonging to the order of Thermoanaerobacterales (2.4% of the total reads), including *Caldanaerobius*, had more than 50% of the cells in the duplication phase (Figure S2).

Unfortunately, it was not possible to calculate the replication index of the other clusters, probably due to the limitation of iRep, which during the analysis discarded the regions with very high and very low coverage, applying a linear regression model relating exclusively to the coverage of the region containing the origin of replication [43].

2.6. Functional Annotation and CAZome Analysis

To assess the metabolic potential of the microbial consortia populating the three sites, the contigs obtained by the assembly were analyzed by Prodigal [44] identifying 14,933 ORFs on Site A, 81,938 ORFs on Site B, and 31,179 ORFs on Site C. Then, the amino acid sequences of the identified ORFs were functionally classified using the COG and SEED databases (Figures 6 and 7).

The analysis of the three samples showed an average comparable distribution of the functional categories reported in both databases. However, by observing in more detail the classification according to the COG database (Figure 6), it is possible to observe a marked difference in relation to the sequences assigned to the functional category "Signal transduction mechanisms" (Category T), where the percentage of ORFs of Site B was two-fold greater than those of Sites A and C. A more

in-depth analysis of the ORFs assigned to this category revealed that this difference was mainly due to the high number of ORF annotated histidine kinases of bacterial origin.

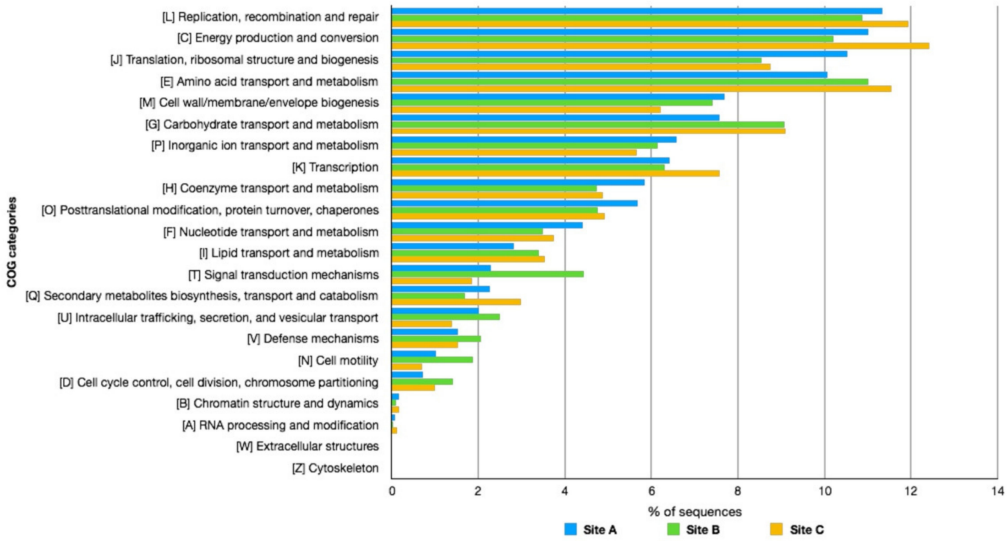


Figure 6. COG analysis of the metagenomes in Pisciarelli hot springs. Sites A, B, and C are compared according to COG functional categories.

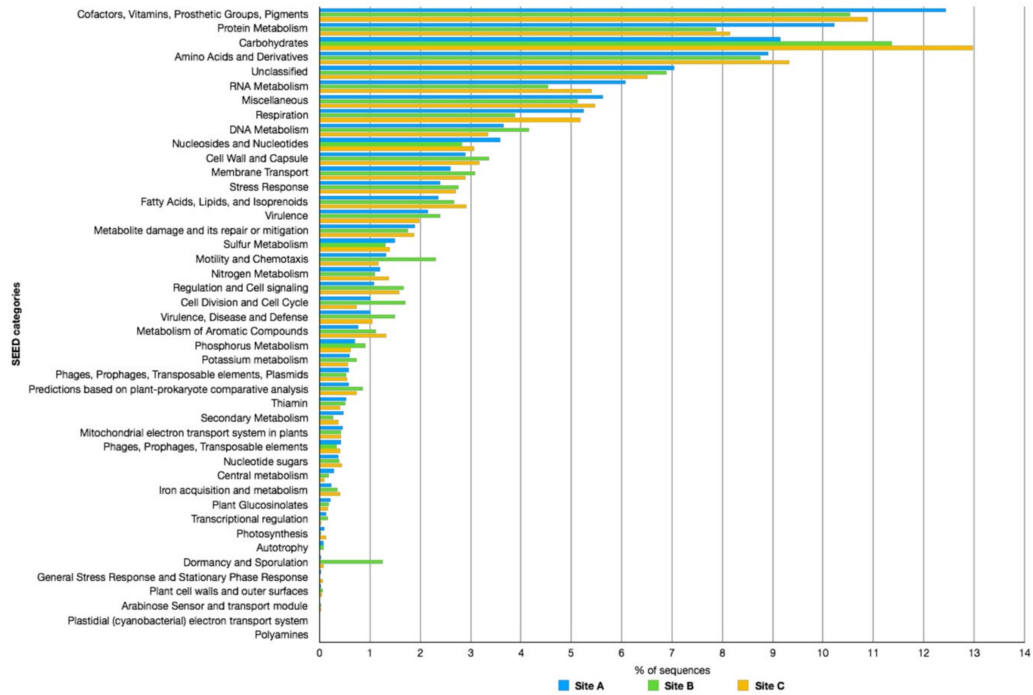


Figure 7. Functional annotation according to the SEED database of the metagenomes in Pisciarelli sites A, B, and C.

In addition, the annotation using the SEED database (Figure 7) showed an average homogeneous distribution between the functional classes, but there were clear differences in relation to the category "Protein Metabolism", more abundant in Site A; "Carbohydrates", more abundant in Site C, where it was also the most represented category; as well as in the categories "Mobility and Chemotaxis" and "Dormancy and Sporulation", in which Site B clearly dominated the other two. In particular, regarding the "Dormancy and Sporulation" category, this was mainly composed of ORFs annotated as Stages 0, I, II, III, IV, and V sporulation proteins, confirming the presence of different sporogenic bacteria in this site.

As for the "Mobility and Chemotaxis" category, the main differences were related to the presence of ORFs of bacterial origin involved in the structure and mobility of the flagellum, while in Site C, the ORFs annotated as "Archaeal Flagellum" and "Bacterial Chemotaxis" (dipeptide-binding ABC transporter) were most abundant.

In all three samples, the largest number of ORFs functionally annotated belonged to the categories "Carbohydrate Transport and Metabolism" and "Carbohydrate" of COG and SEED, respectively. A similar result was previously observed in Pool 1 and Pool 2 and related to the abundant vegetation around the Pisciarelli thermal spring, rich in starch, hemicellulose, and pectins, which could represent an available carbon source for the microbial communities populating these geothermal sites [20].

To map the difference of the enzymatic activities involved in the synthesis, degradation, and modification of carbohydrates (CAZymes) in the Sites A, B, and C, the ORFs of the three samples were analyzed by dbCAN2 [45]. The assessment of the taxonomic origin of the identified CAZymes revealed that in Sites A and C, the highest number of CAZymes belonged to the phylum of the Crenarchaeota (76 and 93% respectively), while 71% of the CAZymes identified in Site B belonged to the phylum of the Firmicutes (Table 6).

Table 6. Taxonomic assignment of CAZymes.

Phylum	Site A	Site B	Site C
Crenarchaeota	76.0	15.2	92.8
Aquificae	16.4	2.9	0
Viruses	3.8	0.6	5.6
Thermodesulfobacteria	1.3	0.2	0
Euryarchaeota	0.6	2.2	0.4
Nitrospirae	0.0	1.7	0
Thermotogae	0.0	2.9	0
Firmicutes	0.0	71.1	0.8
Others (<1%)	1.9	3.2	0.4

The genus analysis (Figure 8) indicated a higher number of *Acidianus*-related CAZymes for Sites A and C. In addition, while Site A had numerous activities related to *Pyrobaculum* and *Desulfurococcus* and to several (hyper)thermophilic bacteria of the phylum Aquificae (*Hydrogenobacter*, *Thermocrinis*, and *Aquifex*), the CAZymes of Site C were mainly assigned to the genera *Metallosphaera, Saccharolobus*, and *Sulfolobus*. Differently, the greater number of CAZymes identified in Site B belonged to the thermophilic bacteria of the genus *Thermoanaerobacterium*, *Thermoanaerobacter*, *Caldanaerobius*, *Caldanaerobacter*, and *Desulfotomaculum* (Figure 8).

However, although it was possible to annotate the identified CAZymes at the genus level, less than 50% of these had an identity ≥95% compared to sequences already present in the Refseq Protein Database, indicating the presence of new sequences related to carbohydrate active enzymes (Figure 9, Tables S4–S6).

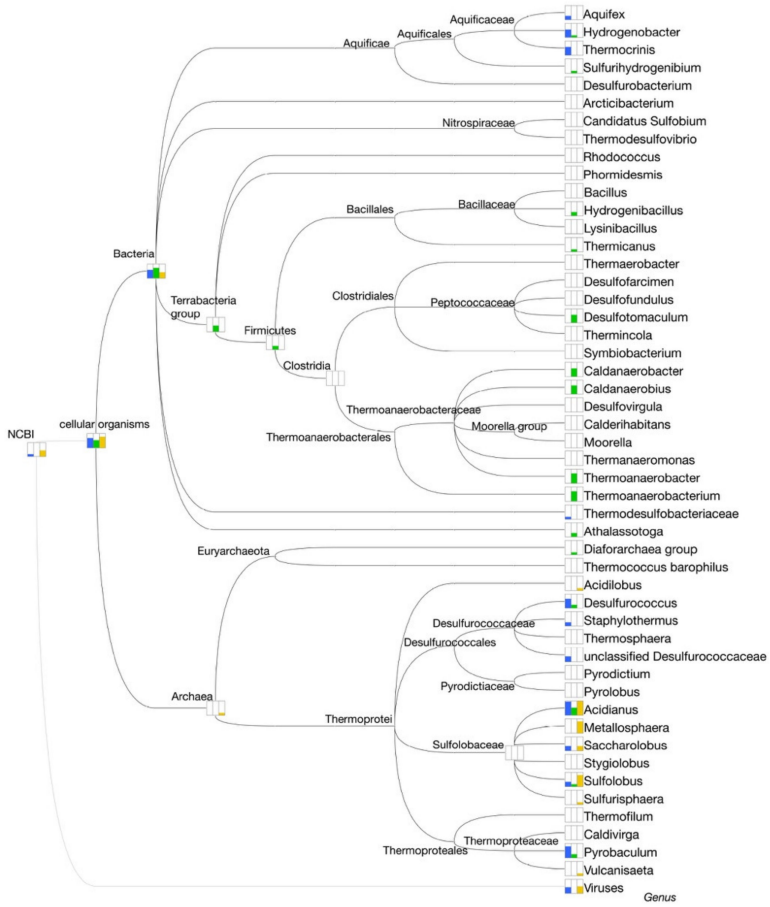


Figure 8. Taxonomic assignment of the CAZymes identified in sites A, B, and C. The filled area of the rectangle indicates, on a logarithmic scale, the number of ORFs assigned to each taxon.

Glycosidases (GHs) represented 28%, 43%, and 27% of the CAZymes identified in Site A, Site B, and Site C, respectively. Among the families of GHs common in all three sites (Figure 10, Table 7), GH13, GH15, GH31, GH57, GH122, and GH133 were identified, which groups enzymes mainly active on α-glycosidic bonds, indicating the presence of pathways of degradation of starch and amylopectins used as energy reserves in plants around the area (mainly ferns, dicots, and grass) [20].

Among the identified GHs, 48 of these were present only in Site B (Table 7). The analysis of aminoacidic sequences showed that these were associated exclusively with (hyper)thermophilic bacteria with identities between 33% and 100% compared to the sequences present in the NCBI NR database (Table S5).

A large fraction of the identified CAZymes was annotated as glycosyltransferase (GTs). In particular, in Sites A and C they represented the most abundant activity class (62% and 58% of the CAZymes, respectively), while in Site B GTs were only 38%.

Eight hypothetical families of GTs were identified exclusively in Site B (Figure 11A), namely GT1, GT13, GT27, GT41, GT76, GT81, and GT104, which are involved in glycosylation mechanisms of proteins and peptides, and GT47, which groups heparan β-glucuronyltransferase, xyloglucan β-galactosyltransferase, heparan synthase, and arabinan α-L-arabinosyltransferase. The aminoacidic sequence of the ORFs classified in these families revealed identity percentages between 48% and 99% compared to the sequences present in NR database, and a bacterial origin related to the genera

Acidithiobacillus, *Anaeromusa*, *Caldanaerobacter*, *Desulfofarcimen*, *Desulfotomaculum*, *Desulfurella*, *Planifilum*, *Syntrophorhabdus*, *Thermoanaerobacter*, *Thermoanaerobacterium*, and *Thiomonas*.

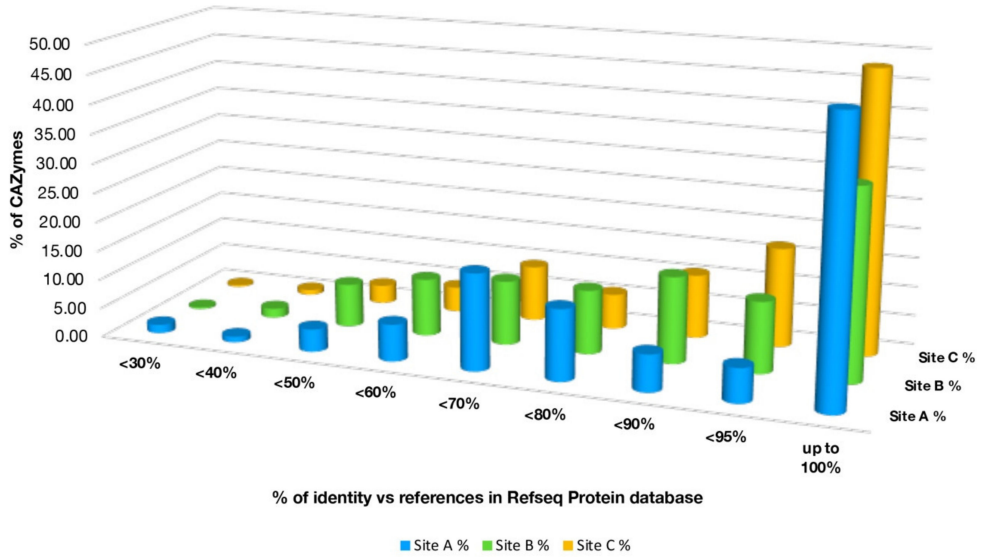


Figure 9. Identity percentages of the CAZymes annotated in the sites A, B, and C against homologs in the Refseq Protein Database.

Table 7. Shared GHs families among the sites.

Sites	Number of GHs Families	Shared GHs Families
Site A, Site B, and Site C	12	GH1, GH12, GH13, GH15, GH31, GH36, GH38, GH57, GH101, GH116, GH122, GH133
Site A and Site B	8	GH4, GH16, GH23, GH26, GH2, GH77, GH99, GH130
Site B and Site C	4	GH3, GH5, GH78, GH109
Site B	48	GH0, GH2, GH6, GH8, GH10, GH14, GH18, GH19, GH20, GH24, GH25, GH27, GH29, GH30, GH32, GH35, GH39, GH42, GH43, GH51, GH52, GH53, GH63, GH65, GH66, GH67, GH73, GH81, GH84, GH92, GH94, GH95, GH97, GH102, GH103, GH105, GH106, GH108, GH120, GH123, GH125, GH127, GH15, GH141, GH144, GH151, GH161, GH163

In addition, with regard to carbohydrate esterases (CEs) (Figure 11B), Site B showed unique families (CE7, CE12, CE15, and CE16) which group putative acetyl xylan esterases, pectin acetyl esterases, rhamnogalacturonan acetyl esterases, 4-O-methyl-glucuronoyl methylesterases, and acetyl-mannan esterases with identity percentages between 58 and 100% with CEs identified in members of the genera *Thermoanaerobacter* and *Desulforella* (Table S5), and which might be involved in the metabolism of hemicellulose polysaccharides.

Site B has also been shown to be particularly rich in hypothetical carbohydrate-binding modules (CBMs), in auxiliary activities (AAs) and in polysaccharide lyases (PLs) (Figure 11C,D). The hypothetical CBMs identified exclusively in Site B are mainly involved in the degradation of starch and amylopectin (CBM20, CBM25, CBM41) and of cellulosic and hemicellulosic polysaccharides (CBM6, CBM22, CBM23, CBM32, CBM54, CBM59) with an identity percentage between 43% and 99% with bacterial sequences,

mostly associated with the *Thermoanaerobacter* and *Caldanaerobacter* genera (Table S5). In contrast, the only exclusive CMB family identified in Site A was CBM4, which groups specific modules for xylan, β-1,3-glucan, β-1,3-1,4-glucan, β-1,6-glucan, and amorphous cellulose but not crystalline cellulose. The single sequence in Site A, annotated as CBM4, shows 97% identity with the cellulase C (GH16) from *Cellvibrio mixus* containing a CBM4 [46].

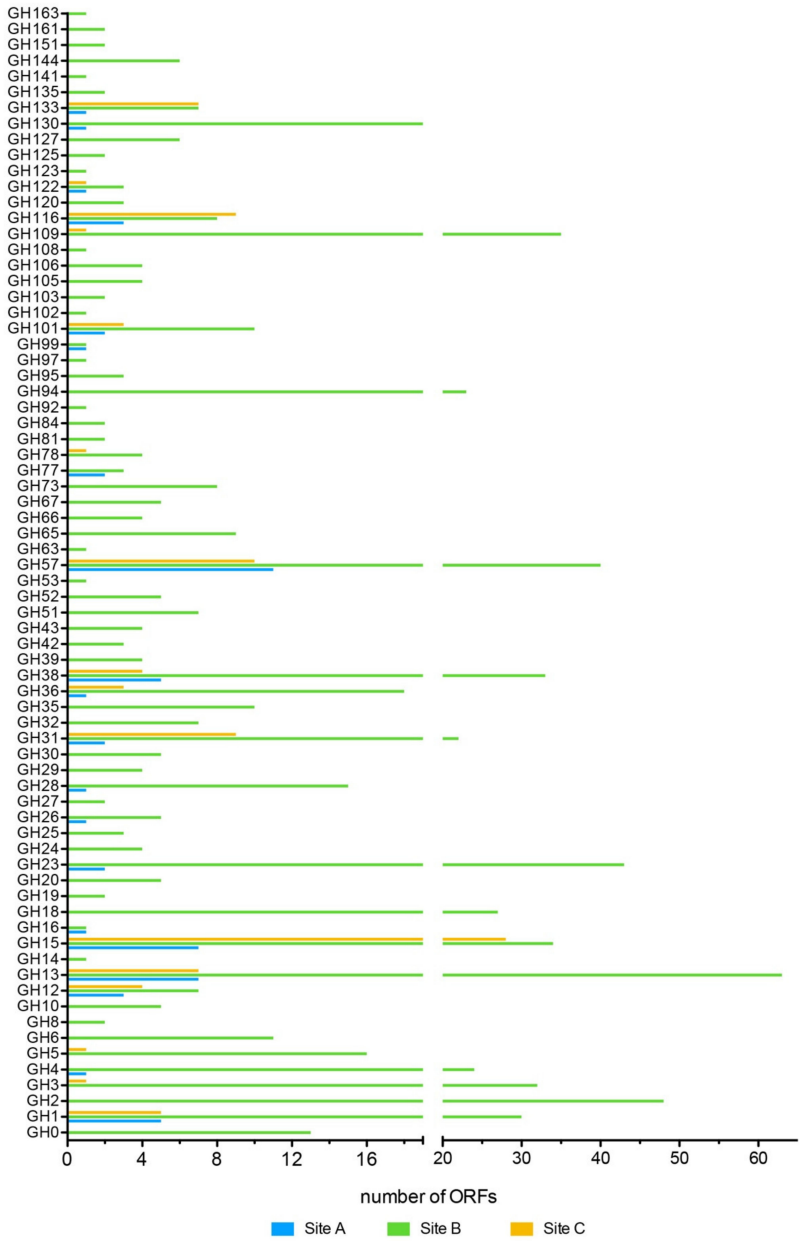


Figure 10. Distribution of glycosidases among the sites A, B, and C. The ORF number assigned to GHs from each sample is displayed.

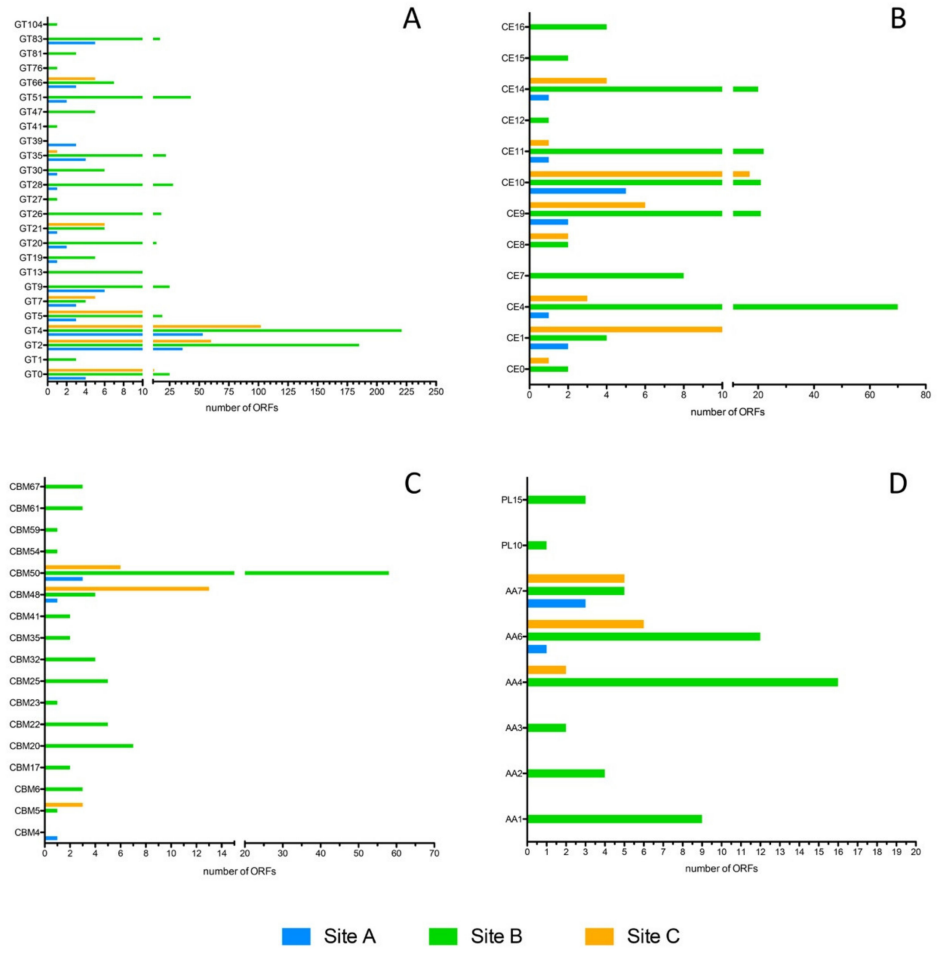


Figure 11. Distribution of the (**A**) glycosyltransferase, (**B**) carbohydrate esterases, (**C**) carbohydrate-binding modules, (**D**) auxiliary activities and polysaccharide lyases among the Pisciarelli sites. The ORFs number assigned to each family from the three samples is displayed.

The hypothetical AAs present in all three sites belong to the families AA6 (1,4-benzoquinone reductases) and AA7 glucooligosaccharide oxidases and chitooligosaccharide oxidases, showing >80% identity in Sites A and C with sequences mainly associated with the phylum of the Crenarchaeota. In Site B, the sequences of which more than 50% of identity have mainly been attributed to the bacterial phyla of Firmicutes and Proteobacteria as well as for the families AA1, AA2, and AA3, were identified exclusively in this sample.

Finally, only in Site B ORFs annotated as PLs were found. Among these, in particular, there was a sequence assigned to the PL10 family, which includes pectate lyase, and three sequences assigned to the PL15 family, which includes alginate lyase, oligoalginate lyase/exo-alginate lyase, heparin lyase I, and heparin lyase III. While the only identified PL10 showed a low identity (47%) with the pectate lyase from *Pelosinus* sp. *UFO1*, the three ORFs classified as PL15 showed identities between 90 and 99% with hypothetical proteins associated with the thermophilic genera *Thermoanaerobacter* and *Caldanaerobacter* (Table S5).

The remarkable number of sequences encoding putative CAZymes makes the Pisciarelli microbial population an attractive source of novel thermophilic biocatalysts for industrial applications.

The functional annotation here described represents a preliminary survey, but already promising a relevant biochemical potential of the microbial consortia of the different geothermal sites in the Pisciarelli area. Indeed, future research will carry out more detailed studies on these extremophilic communities and their CAZomes in order to deepen their knowledge and exploit their biodiversity in biotechnological processes.

3. Materials and Methods

3.1. Pisciarelli Hot Springs Sampling

Samples from the hydrothermal mud/water pools, Site A, Site B, and Site C, respectively, were transferred into sterile tubes, respectively, were closed and immediately transferred to the laboratory for DNA extraction. In situ measurements of temperature and pH were performed by using an HI-93510 thermometer (HANNA instruments, Padova, Italy) equipped with a Pt100 probe, and a pH meter for field use (sensION^{TM} + PH1 equipped with 5051T electrode (HACH)).

3.2. Isolation of DNA

The samples composed by mud/water were centrifuged at 6000× *g* for 20 min at RT, and the sediments were stored at −20 °C. The metagenomic DNA samples were purified from 5 g of sediment collected from each site by following the protocol previously reported [47], except for the lysis step performed by freeze–thawing in dry ice and at 65 °C.

The amounts of obtained metagenomic DNA were quantified by a Qubit 4 Fluorometer using the Qubit™ DNA HS assay kit (Invitrogen-Thermo Fisher Scientific corporation, Waltham, Massachusetts, USA), and DNA quality was independently assessed by visualization on 1% agarose (*w*/*v*) gels.

3.3. Sequencing of mDNA

The extracted and purified mDNA of each site was used for shotgun sequencing with Miseq (Illumina) performed at Novogene Europe, Cambridge (UK), and the datasets obtained were provided as clean reads. The sequencing reads are available in the NCBI Sequence Read Archive (SRA) database under the accession numbers SRR12124857 (Site A), SRR12124856 (Site B), and SRR12124855 (Site C).

The environmental data relative to the Sites A, B, and C (NCBI BioProject PRJNA643424) are available in the Biosamples database under the accession numbers SAMN15414048 (Site A), SAMN15414049 (Site B), and SAMN15414050 (Site C).

3.4. Taxonomic Analysis and Assembly

For microbial diversity analysis, short paired-end Illumina reads (150 bp) were aligned to the nucleotide reference database of NCBI NT by using Blast+/BlastN. The resulting output data of each sample (Site A: reads = 21,090,866 with match in NT; Site B: reads: 18,258,802 with match in NT; Site C: reads: 15'458'279) were used as input for MEGAN6 Community Edition with the following parameters: MinScore = 40.0, MaxExpected = 0.7, TopPercent = 10.0, MinSupportPercent = 0, mode = BlastN) [48].

Clean reads were assembled using MEGAHIT [38] by using min-count = 2 and k-mers 21, 31, 41, 51, 61, 71, 81, 91, and 99.

Obtained contigs ≥1000 bp were grouped into bins by using MyCC [39]; the obtained clusters were validated by CheckM v1.0.12 [40], and reads of each sample were aligned by using Bowtie 2 [42].

3.5. Replicative Estimation

To obtain a replicative estimation, the clusters validated by CheckM were filtered by using an ad hoc pipeline to remove contigs with shorter than 5000 bp, completeness <75%, contamination >2%, and a ratio fragment/Mbp > 175. Again, the reads of each sample were aligned to the remaining clusters, and the SAM files obtained were analyzed by iREP [43].

3.6. Functional Annotation

All the contigs obtained by the assembly procedure were analyzed by using Prodigal [44] to identify the open reading frames. ORFs were analyzed by using Diamond in blastp mode [41] against the NR database and functionally classified by MEGAN6 regarding the SEED and COG databases (MinScore = 35; MaxExpected = 0.01; Top percent = 10; Min support percent = 0.05).

To identify the hypothetical carbohydrate-active enzymes the ORFs were analyzed by the dbCAN2 pipeline [45], and the resulting reads were taxonomically assigned using Diamond in bastp mode, against the NCBI Refseq Protein database.

4. Conclusions

Extremophiles are organisms capable of adapt themselves, survive and thrive in hostile habitats that were previously thought to be adverse or lethal for life [49]. Extreme conditions drive the evolution of their inhabitants, highlighting the role of extremophiles as models for the study of the origin and evolution of life on Earth and provide key insights into the boundaries of life, allowing us to speculate mightily about possible extraterrestrial life forms [50]. Furthermore, the molecular and physiological properties and the remarkable adaptive capabilities of extremophiles make them an attractive source of biocatalysts for diverse applications in biotechnology, biomedicine, and industrial processes.

However, the study of extremophiles is a rather difficult field, mainly constrained by the complexity of reaching their ecological niches and isolating these microbes. Pisciarelli Solfatara hot springs represents a unique ecological niche for the study of hyperthermophiles. This area, so surprisingly dynamic, is affected by sudden geothermal changes such as the increase of the magmatic component of fumaroles, frequent seismic swarms, and bradyseism, indicating that the hydrothermal system undergoes repeated injections of magmatic fluid [22]. These sudden changes are generating hostile environments for survival and growth of (hyper)thermophilic microbial life forms.

The comparative metagenomic study reported here allowed us to understand the complexity of the microbial community in three new sites that were generated by geochemical change happening in this area in July 2019. This study demonstrates that these sites, although very close to each other, showed remarkable differences in terms of pH and temperature that were reflected by significant differences in the microbial consortia inhabiting each site.

Living at such selective pressure might foster the development and retention of a suite of metabolic and physiological adaptations, which could play a key role in ensuring the presence and persistence of life in extreme environments [51].

Indeed, the metagenomic investigation revealed a broad CAZome, correlated to the abundant vegetation present around the Pisciarelli thermal spring, rich in starch, hemicellulose, and pectins, which represent a considerable carbon source for the microorganisms populating the geothermal sites.

The presence of highly sophisticated mechanisms of adaptation together with the availability of specific biochemical pathways sustaining peculiar physiological metabolic capabilities makes the extremophilic microbial communities of Pisciarelli interesting from an astrobiological point of view.

Supplementary Materials: The following are available online. Figure S1. Clustering of metagenomic contigs by MyCC; Figure S2. Estimation of the replication percentages obtained by iRep for cluster 7 of Site A (blue) and clusters 2, 4, and 6 of Site B (green); Table S1. Cluster scores identified by MyCC; Table S2. Cluster validation score by CheckM; Table S3. Stats for the clusters selected for iRep analysis; Table S4. CAZymes annotation in Site A; Table S5. CAZymes annotation in Site B; Table S6. CAZymes annotation in Site C.

Author Contributions: Conceptualization: R.I., A.S., and M.M.; methodology: R.I., F.D.L., and A.S.; investigation: R.I., B.C.-P., F.D.L., N.C., L.M., and A.S.; writing—original draft preparation: R.I., M.M., and A.S.; funding acquisition: B.C.-P. and M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Italian Space Agency for supporting this work co-funding the project "Life in Space: Origin, presence, and persistence of life in space, from molecules to extremophiles" (ASI N. 2019-3-U.0).

Acknowledgments: We thank Rosario Avino and Antonio Carandente of the National Institute of Geophysics and Volcanology (INGV, Naples, Italy) for the technical support during the sampling in the Pisciarelli hot springs.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Saxena, R.; Dhakan, D.B.; Mittal, P.; Waiker, P.; Chowdhury, A.; Ghatak, A.; Sharma, V.K. Metagenomic analysis of hot springs in central india reveals hydrocarbon degrading thermophiles and pathways essential for survival in extreme environments. *Front. Microbiol.* **2016**, *7*, 2123. [CrossRef] [PubMed]
2. Strazzulli, A.; Iacono, R.; Giglio, R.; Moracci, M.; Cobucci-Ponzano, B. Metagenomics of hyperthermophilic environments: Biodiversity and biotechnology. In *Microbial Ecology of Extreme Environments*; Chénard, C., Lauro, F.M., Eds.; Springer International Publishing: Cham, Swizerland, 2017; pp. 103–135. [CrossRef]
3. Cobucci-Ponzano, B.; Rossi, M.; Moracci, M. Interrupted genes in extremophilic archaea: Mechanisms of gene expression in early organisms. *Orig. Life Evol. Biosph.* **2006**, *36*, 487–492. [CrossRef]
4. Onofri, S.; Balucani, N.; Barone, V.; Benedetti, P.; Billi, D.; Balbi, A.; Brucato, J.R.; Cobucci-Ponzano, B.; Costanzo, G.; Rocca, N.; et al. The Italian national project of astrobiology-life in space-origin, presence, persistence of life in space, from molecules to extremophiles. *Astrobiology* **2020**, *20*, 580–582. [CrossRef]
5. Bertrand, J.-C.; Brochier-Armanet, C.; Gouy, M.; Westall, F. For three billion years, microorganisms were the only inhabitants of the earth. In *Environmental Microbiology: Fundamentals and Applications: Microbial Ecology*; Bertrand, J.-C., Caumette, P., Lebaron, P., Matheron, R., Normand, P., Sime-Ngando, T., Eds.; Springer Netherlands: Dordrecht, The Netherlands, 2015; pp. 75–106. [CrossRef]
6. Schulze-Makuch, D. Extremophiles on alien worlds: What types of organismic adaptations are feasible on other planetary bodies. In *Habitability of Other Planets and Satellites*; de Vera, J.-P., Seckbach, J., Eds.; Springer Netherlands: Dordrecht, The Netherlands, 2013; pp. 253–265. [CrossRef]
7. Babu, P.; Chandel, A.K.; Singh, O.V. Therapeutic implications of extremophiles. In *Extremophiles and Their Applications in Medical Processes*; Springer International Publishing: Cham, Swizerland, 2015; pp. 25–35. [CrossRef]
8. Coker, J.A. Extremophiles and biotechnology: Current uses and prospects. *F1000Res* **2016**, *5*. [CrossRef]
9. Durvasula, R.V.; Rao, D.S. Extremophiles: From biology to biotechnology. In *Extremophiles*; Rao, D., Ed.; Taylor & Francis: Boca Raton, FL, USA, 2018; pp. 1–18. [CrossRef]
10. Cobucci-Ponzano, B.; Strazzulli, A.; Iacono, R.; Masturzo, G.; Giglio, R.; Rossi, M.; Moracci, M. Novel thermophilic hemicellulases for the conversion of lignocellulose for second generation biorefineries. *Enzyme Microb. Technol.* **2015**, *78*, 63–73. [CrossRef] [PubMed]
11. Curci, N.; Strazzulli, A.; De Lise, F.; Iacono, R.; Maurelli, L.; Dal Piaz, F.; Cobucci-Ponzano, B.; Moracci, M. Identification of a novel esterase from the thermophilic bacterium Geobacillus thermodenitrificans NG80-2. *Extremophiles* **2019**, *23*, 407–419. [CrossRef] [PubMed]
12. Iacono, R.; Strazzulli, A.; Maurelli, L.; Curci, N.; Casillo, A.; Corsaro, M.M.; Moracci, M.; Cobucci-Ponzano, B. GlcNAc De-*N*-Acetylase from the hyperthermophilic archaeon sulfolobus solfataricus. *Appl. Environ. Microbiol.* **2019**, *85*. [CrossRef]
13. Cobucci-Ponzano, B.; Conte, F.; Benelli, D.; Londei, P.; Flagiello, A.; Monti, M.; Pucci, P.; Rossi, M.; Moracci, M. The gene of an archaeal alpha-L-fucosidase is expressed by translational frameshifting. *Nucleic Acids Res.* **2006**, *34*, 4258–4268. [CrossRef] [PubMed]
14. Strazzulli, A.; Fusco, S.; Cobucci-Ponzano, B.; Moracci, M.; Contursi, P. Metagenomics of microbial and viral life in terrestrial geothermal environments. *Rev. Environ. Sci. Bio.* **2017**, *16*, 425–454. [CrossRef]
15. Khan, M.; Sathya, T.A. Extremozymes from metagenome: Potential applications in food processing. *Crit. Rev. Food Sci. Nutr.* **2018**, *58*, 2017–2025. [CrossRef]
16. Jin, M.; Gai, Y.; Guo, X.; Hou, Y.; Zeng, R. Properties and applications of extremozymes from Deep-Sea extremophilic microorganisms: A mini review. *Mar. Drugs* **2019**, *17*, 656. [CrossRef] [PubMed]
17. Raddadi, N.; Cherif, A.; Daffonchio, D.; Neifar, M.; Fava, F. Biotechnological applications of extremophiles, extremozymes and extremolytes. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 7907–7913. [CrossRef] [PubMed]
18. Menzel, P.; Gudbergsdottir, S.R.; Rike, A.G.; Lin, L.; Zhang, Q.; Contursi, P.; Moracci, M.; Kristjansson, J.K.; Bolduc, B.; Gavrilov, S.; et al. Comparative metagenomics of eight geographically remote terrestrial hot springs. *Microb. Ecol.* **2015**, *70*, 411–424. [CrossRef] [PubMed]
19. Ferrer, M.; Golyshina, O.; Beloqui, A.; Golyshin, P.N. Mining enzymes from extreme environments. *Curr. Opin. Microbiol.* **2007**, *10*, 207–214. [CrossRef]

20. Strazzulli, A.; Cobucci-Ponzano, B.; Iacono, R.; Giglio, R.; Maurelli, L.; Curci, N.; Schiano-di-Cola, C.; Santangelo, A.; Contursi, P.; Lombard, V.; et al. Discovery of hyperstable carbohydrate-active enzymes through metagenomics of extreme environments. *FEBS J.* **2020**, *287*, 1116–1137. [CrossRef]
21. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490–D495. [CrossRef]
22. Chiodini, G.; Caliro, S.; De Martino, P.; Avino, R.; Gherardi, F. Early signals of new volcanic unrest at Campi Flegrei caldera? Insights from geochemical data and physical simulations. *Geology* **2012**, *40*, 943–946. [CrossRef]
23. Chiodini, G.; Caliro, S.; Cardellini, C.; Granieri, D.; Avino, R.; Baldini, A.; Donnini, M.; Minopoli, C. Long-term variations of the Campi Flegrei, Italy, volcanic system as revealed by the monitoring of hydrothermal activity. *J. Geophys. Res. Solid Earth* **2010**, *115*. [CrossRef]
24. Tamburello, G.; Caliro, S.; Chiodini, G.; De Martino, P.; Avino, R.; Minopoli, C.; Carandente, A.; Rouwet, D.; Aiuppa, A.; Costa, A.; et al. Escalating CO_2 degassing at the Pisciarelli fumarolic system, and implications for the ongoing Campi Flegrei unrest. *J. Volcanol. Geoth Res.* **2019**, *384*, 151–157. [CrossRef]
25. Fuchs, T.; Huber, H.; Teiner, K.; Burggraf, S.; Stetter, K.O. Metallosphaera prunae, sp. nov., a novel metal-mobilizing, thermoacidophilic archaeum, isolated from a uranium mine in Germany. *Syst. Appl. Microbiol.* **1995**, *18*, 560–566. [CrossRef]
26. Van Wolferen, M.; Ajon, M.; Driessen, A.J.; Albers, S.V. How hyperthermophiles adapt to change their lives: DNA exchange in extreme conditions. *Extremophiles* **2013**, *17*, 545–563. [CrossRef]
27. White, J.R.; Escobar-Paramo, P.; Mongodin, E.F.; Nelson, K.E.; DiRuggiero, J. Extensive genome rearrangements and multiple horizontal gene transfers in a population of pyrococcus isolates from Vulcano Island, Italy. *Appl. Environ. Microbiol.* **2008**, *74*, 6447–6451. [CrossRef] [PubMed]
28. Ceballos, R.M.; Marceau, C.D.; Marceau, J.O.; Morris, S.; Clore, A.J.; Stedman, K.M. Differential virus host-ranges of the Fuselloviridae of hyperthermophilic Archaea: Implications for evolution in extreme environments. *Front. Microbiol.* **2012**, *3*, 295. [CrossRef] [PubMed]
29. Panda, S.K.; Jyoti, V.; Bhadra, B.; Nayak, K.C.; Shivaji, S.; Rainey, F.A.; Das, S.K. Thiomonas bhubaneswarensis sp. nov., an obligately mixotrophic, moderately thermophilic, thiosulfate-oxidizing bacterium. *Int. J. Syst. Evol. Microbiol.* **2009**, *59*, 2171–2175. [CrossRef] [PubMed]
30. Vesteinsdottir, H.; Reynisdottir, D.B.; Orlygsson, J. *Thiomonas islandica* sp. nov., a moderately thermophilic, hydrogen- and sulfur-oxidizing betaproteobacterium isolated from a hot spring. *Int. J. Syst. Evol. Microbiol.* **2011**, *61*, 132–137. [CrossRef]
31. Asano, R.; Hirooka, K.; Nakai, Y. Middle-thermophilic sulfur-oxidizing bacteria Thiomonas sp RAN5 strain for hydrogen sulfide removal. *J. Air Waste Manag.* **2012**, *62*, 38–43. [CrossRef]
32. Hallberg, K.B.; Lindstrom, E.B. Characterization of Thiobacillus caldus sp. nov., a moderately thermophilic acidophile. *Microbiology* **1994**, *140*, 3451–3456. [CrossRef]
33. Etcheverría, A.I.; Lucchesi, P.M.A.; Krüger, A.; Bentancor, A.B.; Padola, N.L. Escherichia coli in animals. In *Escherichia Coli in the Americas*; Torres, A.G., Ed.; Springer International Publishing: Cham, Swizerland, 2016; pp. 149–172. [CrossRef]
34. Samanta, I.; Bandyopadhyay, S. Clostridium. In *Antimicrobial Resistance in Agriculture*; Indranil Samanta, S.B., Ed.; Academic Press: Cambridge, MA, USA, 2020; pp. 253–262. [CrossRef]
35. Chan, C.S.; Chan, K.G.; Tay, Y.L.; Chua, Y.H.; Goh, K.M. Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. *Front. Microbiol.* **2015**, *6*, 177. [CrossRef]
36. Ghilamicael, A.M.; Boga, H.I.; Anami, S.E.; Mehari, T.; Budambula, N.L.M. Potential human pathogenic bacteria in five hot springs in Eritrea revealed by next generation sequencing. *PLoS ONE* **2018**, *13*, e0194554. [CrossRef]
37. Liu, L.; Jiao, J.-Y.; Fang, B.-Z.; Lv, A.-P.; Ming, Y.-Z.; Li, M.-M.; Salam, N.; Li, W.-J. Isolation of clostridium from Yunnan-Tibet hot springs and description of clostridium thermarum sp. nov. with lignocellulosic ethanol production. *Syst. Appl. Microbiol.* **2020**, *43*. [CrossRef]
38. Li, D.; Liu, C.M.; Luo, R.; Sadakane, K.; Lam, T.W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **2015**, *31*, 1674–1676. [CrossRef] [PubMed]
39. Lin, H.H.; Liao, Y.C. Accurate binning of metagenomic contigs via automated clustering sequences using information of genomic signatures and marker genes. *Sci. Rep.* **2016**, *6*, 24175. [CrossRef] [PubMed]

40. 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.* **2015**, *25*, 1043–1055. [CrossRef] [PubMed]
41. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60. [CrossRef] [PubMed]
42. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]
43. Brown, C.T.; Olm, M.R.; Thomas, B.C.; Banfield, J.F. Measurement of bacterial replication rates in microbial communities. *Nat. Biotechnol.* **2016**, *34*, 1256–1263. [CrossRef]
44. Hyatt, D.; Chen, G.L.; Locascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **2010**, *11*, 119. [CrossRef]
45. Zhang, H.; Yohe, T.; Huang, L.; Entwistle, S.; Wu, P.; Yang, Z.; Busk, P.K.; Xu, Y.; Yin, Y. dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **2018**, *46*, W95–W101. [CrossRef]
46. Centeno, M.S.; Goyal, A.; Prates, J.A.; Ferreira, L.M.; Gilbert, H.J.; Fontes, C.M. Novel modular enzymes encoded by a cellulase gene cluster in Cellvibrio mixtus. *FEMS Microbiol. Lett.* **2006**, *265*, 26–34. [CrossRef]
47. Zhou, J.; Bruns, M.A.; Tiedje, J.M. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **1996**, *62*, 316–322. [CrossRef]
48. Huson, D.H.; Beier, S.; Flade, I.; Gorska, A.; El-Hadidi, M.; Mitra, S.; Ruscheweyh, H.J.; Tappu, R. MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* **2016**, *12*, e1004957. [CrossRef] [PubMed]
49. Orellana, R.; Macaya, C.; Bravo, G.; Dorochesi, F.; Cumsille, A.; Valencia, R.; Rojas, C.; Seeger, M. Living at the frontiers of life: Extremophiles in Chile and their potential for bioremediation. *Front. Microbiol.* **2018**, *9*, 2309. [CrossRef] [PubMed]
50. Merino, N.; Aronson, H.S.; Bojanova, D.P.; Feyhl-Buska, J.; Wong, M.L.; Zhang, S.; Giovannelli, D. Living at the extremes: Extremophiles and the limits of life in a planetary context. *Front. Microbiol.* **2019**, *10*, 780. [CrossRef] [PubMed]
51. Dong, Y.; Sanford, R.A.; Inskeep, W.P.; Srivastava, V.; Bulone, V.; Fields, C.J.; Yau, P.M.; Sivaguru, M.; Ahren, D.; Fouke, K.W.; et al. Physiology, metabolism, and fossilization of Hot-Spring filamentous microbial mats. *Astrobiology* **2019**, *19*, 1442–1458. [CrossRef]

Sample Availability: Samples of the collected sediments are available from the authors.

molecules

MDPI

Article

Transcript Regulation of the Recoded Archaeal α-L-Fucosidase In Vivo

Federica De Lise [1], Roberta Iacono [1,2], Andrea Strazzulli [2,3], Rosa Giglio [1,†], Nicola Curci [1,2], Luisa Maurelli [1], Rosario Avino [4], Antonio Carandente [4], Stefano Caliro [4], Alessandra Tortora [5], Fabio Lorenzini [5], Paola Di Donato [6,7], Marco Moracci [1,2,3] and Beatrice Cobucci-Ponzano [1,*]

[1] Institute of Biosciences and BioResources, National Research Council of Italy, Via P. Castellino 111, 80131 Naples, Italy; federica.delise@ibbr.cnr.it (F.D.L.); roberta.iacono@unina.it (R.I.); Rosa.Giglio@savioindustrial.it (R.G.); nicola.curci@unina.it (N.C.); luisa.maurelli@ibbr.cnr.it (L.M.); marco.moracci@unina.it (M.M.)
[2] Department of Biology, University of Naples "Federico II", Complesso Universitario Di Monte S. Angelo, Via Cupa Nuova Cinthia 21, 80126 Naples, Italy; andrea.strazzulli@unina.it
[3] Task Force on Microbiome Studies, University of Naples Federico II, 80134 Naples, Italy
[4] National Institute of Geophysics and Volcanology, Via Diocleziano, 328, 80125 Naples, Italy; rosario.avino@ingv.it (R.A.); antonio.carandente@ingv.it (A.C.); stefano.caliro@ingv.it (S.C.)
[5] Kayser Italia Srl. Via di Popogna, 501, 57128 Livorno, Italy; a.tortora@kayser.it (A.T.); f.lorenzini@kayser.it (F.L.)
[6] Institute of Biomolecular Chemistry, National Research Council of Italy, Via Campi Flegrei 34, Pozzuoli, 80078 Naples, Italy; pdidonato@uniparthenope.it
[7] Department of Science and Technology, University of Naples "Parthenope", Centro Direzionale Isola C4, 80143 Naples, Italy
* Correspondence: beatrice.cobucciponzano@ibbr.cnr.it.com
† Current Address: Savio Industrial Srl., Via Emilia 21, 27100 Pavia, Italy.

Citation: De Lise, F.; Iacono, R.; Strazzulli, A.; Giglio, R.; Curci, N.; Maurelli, L.; Avino, R.; Carandente, A.; Caliro, S.; Tortora, A.; et al. Transcript Regulation of the Recoded Archaeal α-L-Fucosidase In Vivo. *Molecules* **2021**, *26*, 1861. https://doi.org/10.3390/molecules26071861

Academic Editors: Raffaele Saladino and Leonidas A. Phylactou

Received: 7 December 2020
Accepted: 19 March 2021
Published: 25 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: Genetic decoding is flexible, due to programmed deviation of the ribosomes from standard translational rules, globally termed "recoding". In *Archaea*, recoding has been unequivocally determined only for termination codon readthrough events that regulate the incorporation of the unusual amino acids selenocysteine and pyrrolysine, and for −1 programmed frameshifting that allow the expression of a fully functional α-L-fucosidase in the crenarchaeon *Saccharolobus solfataricus*, in which several functional interrupted genes have been identified. Increasing evidence suggests that the flexibility of the genetic code decoding could provide an evolutionary advantage in extreme conditions, therefore, the identification and study of interrupted genes in extremophilic Archaea could be important from an astrobiological point of view, providing new information on the origin and evolution of the genetic code and on the limits of life on Earth. In order to shed some light on the mechanism of programmed −1 frameshifting in Archaea, here we report, for the first time, on the analysis of the transcription of this recoded archaeal α-L-fucosidase and of its full-length mutant in different growth conditions in vivo. We found that only the wild type mRNA significantly increased in *S. solfataricus* after cold shock and in cells grown in minimal medium containing hydrolyzed xyloglucan as carbon source. Our results indicated that the increased level of *fucA* mRNA cannot be explained by transcript up-regulation alone. A different mechanism related to translation efficiency is discussed.

Keywords: *Archaea*; extremophiles; recoding; programmed frameshifting; limits of life

1. Introduction

The decoding of genetic information into polypeptides is a dynamic mechanism in which the standard rules of decoding can be altered in special cases. In fact, in particular genes, signals encoded in the mRNA reprogram the ribosome to read the message in an alternative way, a phenomenon called translational recoding [1]. Translational recoding has been identified in organisms from all three domains of life and in viruses, and an updated list of the genes regulated by this mechanism can be found in the Recode[2] database [2]. Recoding has crucial roles in the regulation of gene expression and the most common events are stop codon readthrough and programmed frameshifting (PRF) [1,3–7]. In stop codon readthrough a stop codon is decoded as a sense codon by a near-cognate tRNA. In addition to readthrough by near-cognate aa-tRNAs, stop codons can be recorded by the specialized tRNAs with an anticodon that is complementary to the stop codon, such as $tRNA^{Pyl}$ or $tRNA^{Sec}$, encoding for the unusual amino acids pyrrolysine [8] and selenocysteine [9]. Specific stimulatory elements downstream to the stop codon regulate this process. Events of stop codon readthrough have been reported in all the three domains of life [6]; in fact, a recent study showed that vascular endothelial growth factor-A mRNA in mammalian endothelial cells undergoes programmed translational readthrough generating an isoform containing a C-terminus extension [10]. Moreover, recoding is of interest for biomedical applications. Recent studies revealed that the premature termination codons suppression by specific drugs named "readthrough agents" may play a role in the clinical treatment of genetic diseases caused by nonsense mutations such as cystic fibrosis and Duchenne muscular dystrophy [11]. In PRF, ribosomes are induced to shift to an alternative, overlapping reading frame 1 nt 3'-wards (+1 frameshifting) or 5'-wards (−1 frameshifting) of the mRNA. The frequency of this process varies in different genes where it is under the control of sophisticated mechanisms [1]. The PRF has been studied extensively in viruses, retrotransposons and insertion elements for which many cases are documented [12,13]. Among cellular genes, where this phenomenon is less common, the best studied is the Antizyme expressed by a +1 PRF, from yeast and protists up to humans, that functions both as a sensor of the polyamine levels and as an effector of a self-regulating circuit [14].

In Archaea, recoding, which was deeply studied only recently, was unequivocally demonstrated only for termination codon readthrough events that regulate the incorporation of the unusual amino acids selenocysteine and pyrrolysine [8,15], and −1 PRF that allow the expression of a fully functional α-L-fucosidase in the crenarchaeon *Saccharolobus solfataricus* [16–21]. This gene, named *fucA*, is organized in two open reading frames (ORFs) SSO11867 and SSO3060 of 81 and 426 amino acids, respectively, which are separated by a −1 frameshifting in a 40 bases overlap (Figure 1A). The analysis of the region of overlap between the two ORFs showed the characteristic features of the genes expressed by −1 PRF, including a heptanucleotide A-AAA-AAT (codons are shown in the zero frame) named slippery sequence, where the −1 PRF of the ribosome takes place, flanked by two rare CAC codons in tandem, and a putative stem–loop secondary structure (Figure 1A) resembling, respectively, the bacterial Shine–Dalgarno-like sites and stem–loops/hairpins, both with the function to slow down the translating ribosomes and promote −1 PRF. Remarkably, we demonstrated that a full-length mutant of gene, named *framefucA*, obtained by inserting specific site-directed mutations in the *fucA* gene in the positions that were predicted to generate by −1 PRF a complete polypeptide (Figure 1B) led to a functional enzyme α-L-fucosidase, named Ssα-fuc, of 495 amino acids, which resulted in it being thermophilic, thermostable, and having an unusual nonameric structure [16–19,22]. In addition, we showed that *fucA* is expressed by −1 PRF in both *E. coli* and *S. solfataricus* demonstrating, for the first time, that this kind of recoding is present in Archaea [20]. To date, only 8 archaeal α-L-fucosidases are reported and that from *S. solfataricus* is the only one characterized. It is interesting to note that the *S. solfataricus* strains P2 and 98/2, although isolated in the Pisciarelli solfataric field in Italy and in the Yellowstone National Park, respectively, shows the same interruption, suggesting a

conserved regulating mechanism for this gene organization. More recently, −1 frameshifting also appears to be used by the siphoviruses tailed virus 1 (HVTV-1) and three viruses (HCTV-1, 2 and 5) that infect halophilic archaea, although the one used by the haloarchaealmyovirus tailed virus 2 (HSTV-2) is likely +1 frameshifting [23,24]. In addition, frameshifting is likely involved in the synthesis of magnesium chelatase from the archaea *Methanocaldococcus* and *Methanococcus* [25]. However, no detailed studies on the regulatory mechanism of these genes are reported. In Archaea, several functional interrupted genes have been identified in *S. solfataricus* [26]. Increasing evidence suggests that the flexibility of the genetic code decoding is a trait selected during evolution that may increase microbial fitness under certain conditions [27]. This could be particularly relevant in extreme environments, which, contrary to common believe, are not immutable but subjected to sudden changes that greatly, and temporarily, modify the chemical-physical parameters and to which microorganisms must adapt. For these reasons the identification and study of interrupted genes in extremophilic Archaea are important from an astrobiological point of view and can provide new information on the origin and evolution of the genetic code and on the limits of life on Earth and beyond.

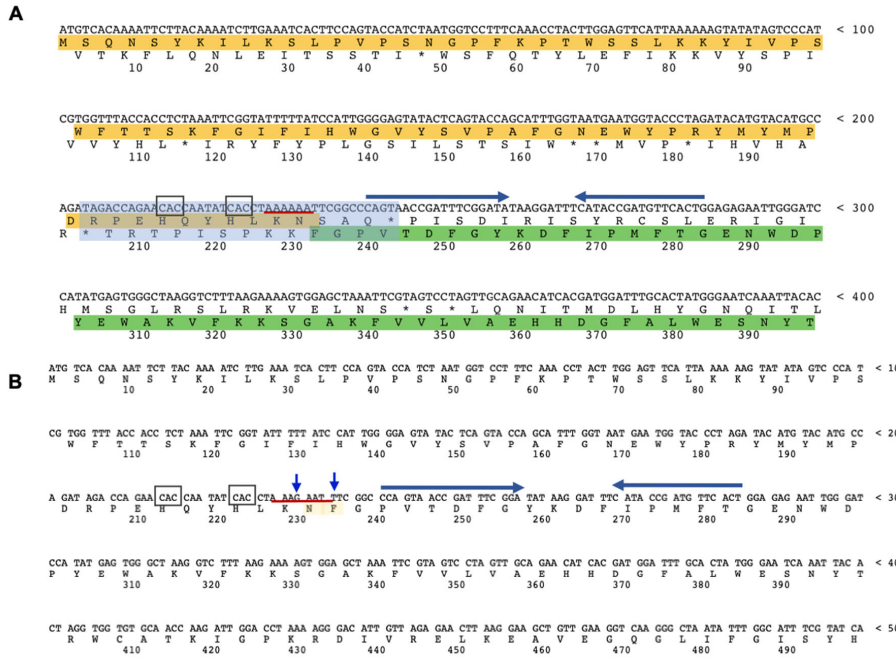


Figure 1. The α-L-fucosidase gene. (**A**) *fucA* gene sequence. The N-terminal SSO11867 ORF (highlighted in yellow) is in the zero frame, the C-terminal SSO3060 ORF (highlighted in green), for which only a fragment is shown, is in the −1 frame. The 40 bp region of overlap between the two ORFs is highlighted with a light blue rectangle. The slippery heptameric sequence is underlined with a red line; the rare codons are boxed, and the arrows indicate the stems of the putative mRNA secondary structure. (**B**) *framefucA* mutant gene (only a fragment is shown). The blue arrows indicate the mutated nucleotides in the slippery sequence, the insertion allowing to restore a single frame between the two ORFs (involved amino acids highlighted in yellow).

Unfortunately, the study of recoding is still very limited in extremophilic Archaea due to the difficulties of growing them in laboratory, the lack of reliable tools for gene manipulation, and for the limited knowledge of the physiology of these organisms in vivo. *fucA* is the only archaeal gene that has been demonstrated to be expressed by −1 PRF and,

since it encodes for an enzyme that can be easily assayed in-vitro, it is an ideal molecular system to study −1PRF in vivo. In particular, understanding why *fucA*, a gene presumably involved in carbohydrate metabolism, is expressed by recoding and if its expression is regulated by specific growth conditions or metabolites, could help to shed some light on how this mechanism evolved and is regulated in Archaea. Here, to test if the expression of *fucA* is regulated in vivo, and in an effort of identifying external effectors involved in its expression, we analyzed its transcription and the enzymatic activity of the α-L-fucosidase in different growth conditions. In particular, we compared the wild type strain, in which the expression of *fucA* is controlled by −1 PRF, with a mutant strain in which we inserted the full-length gene and whose expression is therefore not translationally regulated by −1PRF. We report here that, in some conditions, the mRNA level of the wild type transcript increased up to 10-fold, while the level of the full-length transcript remained almost unchanged. The possibility that the increase in *fucA* transcript in the wild type is due to improved translation efficiency rather than transcription up-regulation is discussed.

2. Results

2.1. Knocked-Out and Full-Length fucA S. solfataricus Strains

The essentiality of the α-L-fucosidase gene in *S. solfataricus* was analyzed by preparing a deletion mutant PBL2025:ΔSSO3060-SSO11867 (Del) and comparing it with the parental wild type *(WT)* and PBL2025 (a *S. solfataricus* strain deleted of 50 genes, from ORFs SSO3004 to SSO3050, many of which encode for carbohydrate-active enzymes) strains in standard conditions. To prepare the Del mutant, 444 nucleotides, internal to the gene, from position 192 to position 638, were deleted. The strain has been controlled by PCR. The deletion resulted in the introduction of a stop codon after 213 nucleotides from the ATG of the first ORF SSO11867. This could only result in the translation of a polypeptide of 71 amino acids, ruling out the translation of a full-length protein and a functional enzyme. As reported in Figure 2A, the Del strain is viable, clearly indicating that *fucA* is not an essential gene for *S. solfataricus* grown in this condition. In order to analyse the effect of the presence of a full-length gene, not regulated by −1PRF, we prepared a mutant strain of *S. solfataricus* in which the interrupted wild type *fucA* gene was substituted with the full-length mutant *framefucA* (FFuc strain). This mutant has been prepared by replacing the same 444 nucleotide sequence of the wild type, as described above, with those of the *framefucA* full-length mutant (Figure 1B). Therefore, in all these mutants the possible transcription regulatory signals remained unchanged. By comparing the growth curves, we observed that all strains were viable, but the mutants had a slightly longer latency phase than *WT* (Figure 2A). Western Blot analysis (Figure 2B) performed on the cellular extracts of both wild type and the two mutants, using antibodies against α-L-fucosidase, confirmed that the higher molecular band revealed in *WT* cellular extracts corresponded to the oligomeric form of the α-L-fucosidase as previously reported [20,22]. As expected, a more intense signal was observed in the full-length mutant FFuc strain. The high apparent molecular weight of the bands in lanes 3–4 is due to the higher stability of the nonameric structure of the enzyme in the *S. solfataricus* extract if compared to the purified recombinant α-L-fucosidase, as already reported [20,22]. As reported previously [20], unspecific signals for bands lower than 97 kDa and a very faint band visible in the Del lane, were detected. The band observed at about 70 kDa in the Del mutant probably represent a multimer of the 71 amino acids polypeptide.

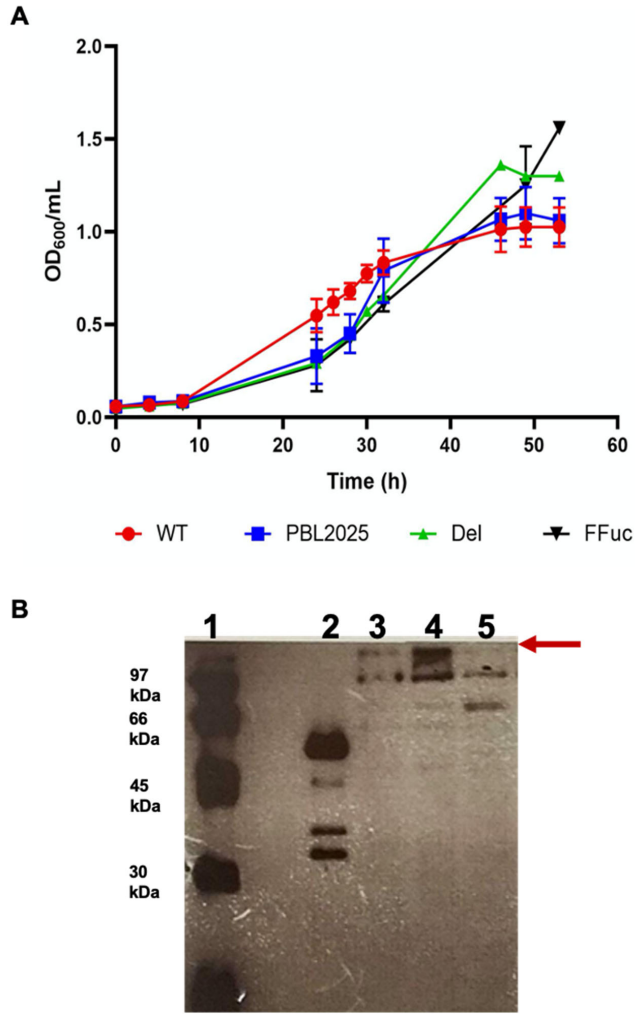


Figure 2. *S. solfataricus* wild type and mutant strains. (**A**) Representative growth curves *of S. solfataricus* wild type (WT), PBL2025, Del and FFuc strains (red, blue, green and black lines, respectively) in YCS medium at 80 °C. Error bars reported represent standard deviations. (**B**) Western blot analysis, performed with anti-α-L-fucosidase antibody, on cellular extracts of *WT*, FFuc and Del mutants. Lane 1: ECL markers (97-66-45-30 kDa), lane 2: α-L-fucosidase 0.1 mg/mL; the additional bands observed are due to partial proteolysis as already reported [22]; lane 3: wild type extract; lane 4: FFuc extract; lane 5: Del extract. The apparent high molecular bands corresponding to the α-fucosidase in vivo are indicated by a red arrow.

WT and FFuc strains grown in YCS were analyzed after 48 h of growth (late exponential phase, FFuc 1.0, *WT* 0.9 OD_{600}) by contrast phase microscopy and no clear morphological differences could be observed between the two strains (Figure 3A). To get more insights on the possible differences between the two strains at molecular level, we analyzed and compared the transcript level of *fucA* and *framefucA,* and the enzymatic units of α-L-fucosidase in *WT* and FFuc strains. In particular, the activity of the enzyme is a convenient indication of the expression of a full-length polypeptide. Interestingly, in the FFuc mutant the α-L-fucosidase activity was 8-fold higher than in the *WT* strain (18.0 vs. 2.3

mU/mg) in standard conditions, as expected from a full-length gene. The mRNAs from *WT* and FFuc strains, grown in standard conditions and recovered in the late exponential phase as above, were analyzed by Real-time PCR. The ratio between wild type and mutant mRNA was the result of three different measures and the transcript level of each sample was normalized by using 16S rRNA specific oligonucleotides. The analyses showed that the amount of mRNA extracted from the FFuc mutant (bearing the full-length gene) is 100-fold higher than the mRNA extracted from the wild type strain in which the gene was interrupted (Figure 3B). It is well known that mRNAs with a premature stop codon (PTC) are recognized and targeted for degradation [28–30]. Thus, this result suggested that the presence of the −1 frameshift could acted as a *cis-acting* mRNA destabilizing element, targeting part of the wild type mRNA for degradation. By contrast, the full-length mRNA remained stable.

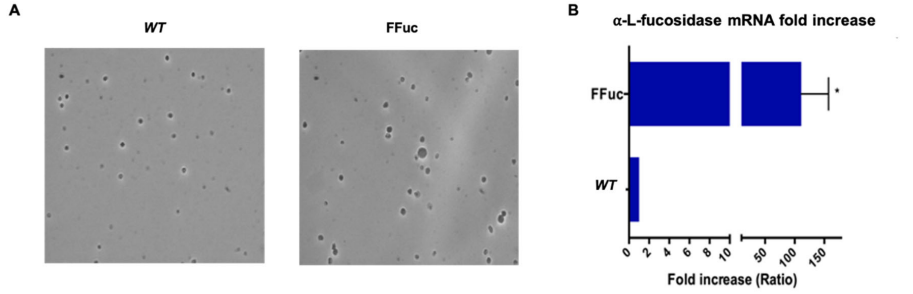


Figure 3. Comparison of *S. solfataricus* wild type and FFuc strains. (**A**) Contrast phase microscopy of of *WT* and FFuc cell cultures harvested in late exponential phase; (**B**) Transcriptional analysis of *fucA* and *framefucA* mRNA in *S. solfataricus WT* and FFuc strains, respectively, analyzed by real-time PCR. The amount of *fucA* and *framefucA* mRNA were determined in cultured cells and normalized against 16S (internal control). The amount of mRNA in the *WT* was set as 1 and that present in FFuc was normalized accordingly. Data are representative of three measurements and are expressed as the mean ± SD. Statistical significance was performed using the two-tailed paired Student's *t* test (* $p < 0.05$).

2.2. Environmental Conditions Variation in Pisciarelli Solfatara

It has already been well documented that PRF increases the coding potential of the genome, it is often used to ensure a defined stoichiometry of protein products, to expand the variability of cellular proteomes or adapt to changing environments [27]. The special living conditions of Archaea, in particular of *S. solfataricus* growing at $T = 80$ °C and pH 3.0–5.0 in geothermal sites, make this organism an interesting model system to study whether translational recoding is involved in adaptations to extreme environment, stress responses and changing growth conditions. On the other hand, to predict which specific elements could be involved in the regulation of *fucA* is not easy. To evaluate the effect of the natural environment we exposed the wild type laboratory strain to a water-mud pool in the Pisciarelli solfataric field (Figure 6), where it was originally isolated [31]. The *S. solfataricus* grown in the laboratory up to 0.17 OD_{600nm} was divided into 3 groups of sample cultures incubated as follows: (i) in the Pisciarelli solfatara using tubes capped with a 0.22 μm filter, which allowed the exchange of trace elements present in the solfatara water, but not of microorganisms and sediments; (ii) in the lab in the same tubes, but in a shaker at 80 °C; (iii) in the lab in a standard laboratory flask at 80 °C, in a shaker. To incubate the samples in the Pisciarelli solfatara, a device, specially designed and built in collaboration with Kayser Italia, was used. The device was anchored and kept floating by a buoy (Figure 4A). The temperature of the Pisciarelli solfatara pool was monitored using a thermometer inserted in the incubation device. The OD_{600nm} were monitored in the laboratory controls for the same incubation time as in Pisciarelli, and for the following three days. As shown in Figure 4B, the growth of the controls in the tubes was slower than in the flask, probably due to reduced oxygenation. When the control culture in the flask exceeded 1.0 OD_{600nm},

the tubes were recovered from the solfataric pool. At the end of incubation, the following OD_{600nm} were measured: Pisciarelli tubes 0.21, lab flask 1.18 and lab tubes 0.42. During the experiment, the temperature in Pisciarelli was rather stable, between 80 and 90 °C; however, fluctuations were also observed with a negative peak at 66 °C, and increments up to 96 °C, in response to heavy rain events and/or changing in the hydrothermal gas flux discharged by the Pisciarelli pool (Figure 4D). A morphological analysis of the three samples by contrast phase optical microscopy revealed the presence of cellular aggregates only in the samples incubated in the Pisciarelli solfatara (Figure 4C). This is similar to UV-induced stress response. In fact, it has been reported that the stress induced by UV irradiation of *S. solfataricus* induces an archaeal pili system, which mediates cellular aggregation in response to UV damage [32,33].

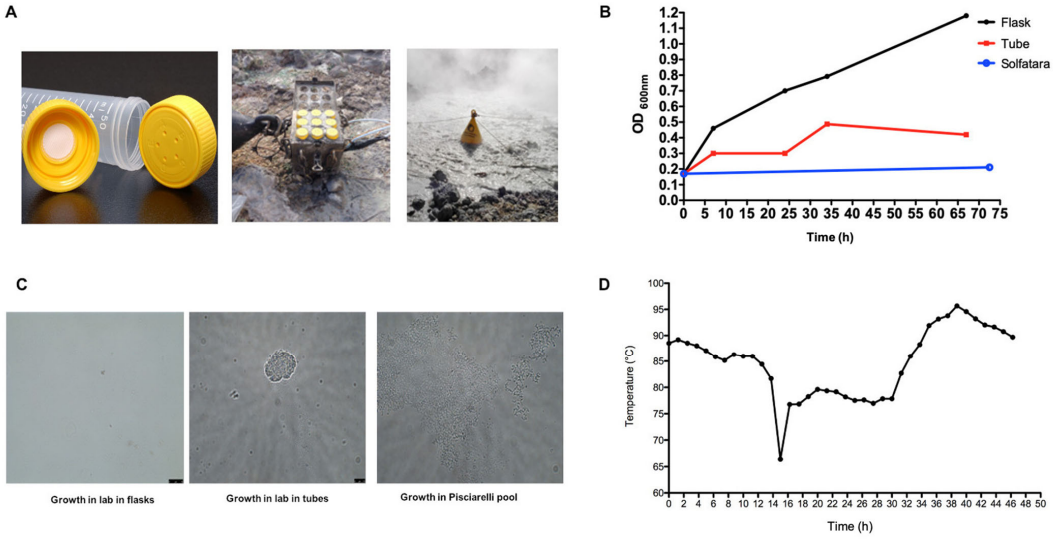


Figure 4. Growth of *S. solfataricus* in solfatara Pisciarelli. (**A**) Device for the *S. solfataricus* growth in the Pisciarelli pool; (**B**) growth curves of *S. solfataricus WT* in YCS medium, in lab (flasks and plastic tubes at 80 °C, black and red, respectively) and in Pisciarelli pool in plastic tubes (blue); (**C**) contrast phase microscopy of cultures in lab, in flasks and plastic tubes, respectively and in Pisciarelli pool in plastic tubes; (**D**) fluctuation temperatures measured in Pisciarelli pool during *S. solfataricus* incubation.

Unfortunately, the very low amount of proteins extracted from the sample incubated in the Pisciarelli pool (1.6 mg/mL), if compared to the control samples in the tubes (10.9 mg/mL) and in the flask (16.2 mg/mL), prevented a detailed comparative proteomic analysis of the three samples in order to evaluate *fucA* and other interrupted genes expression in these conditions. Since the incubation of samples in the solfataric field was difficult to replicate and monitor, we decided to analyse whether different growth conditions could regulate the transcription and translation of *fucA* under controlled experiments in the lab.

2.3. Transcript and Enzymatic Activity Levels of α-L-Fucosidase in Different Conditions

The reason why the expression of *fucA* is regulated by frameshifting in vivo is not known. Several lines of evidence allowed us to exclude that −1 PRF is used to set the ratio of two polypeptides for the α-L-fucosidase but, rather, we suggested that it is used to regulate the expression of a functional full-length product [17–20]. Thus, the natural frameshifting levels of *fucA* could likely vary depending on growth condition and physiological state, as reported for *E. coli* in which frameshifting levels increase entering in the

stationary phase, presumably due to starvation and/or aa-tRNAs limitation [34,35], or in the case in which a higher α-L-fucosidase activity is requested for specific physiological reason.

We observed that, in standard growth conditions, *fucA* produced a rare transcript [20] and expressed a low α-L-fucosidase activity. To get more insights into the possible regulation of the transcription and translation of *fucA*, the *WT* and FFuc strain were grown in different conditions and compared at molecular level by measuring transcript level, by Real-time PCR, and the α-L-fucosidase activity, by enzymatic assays.

2.3.1. Cold Shock and UV Irradiation

As observed above, in its natural environment, *S. solfataricus* may have to face sudden changes in temperature to which it must quickly respond. To evaluate the possible impact of cold shock on the *fucA* gene expression, cold shock time course experiments were carried out and growth curves of the wild type and FFuc strains were monitored up to stationary phase (Figure 5). As reported in Figure 5A,B, cells viability of both strains at 65 °C, and after cold shock at 4 °C, was not affected, and the growth curves were comparable to the control at 80 °C, with no significant differences. Instead, *fucA* mRNA showed a 10-fold increase in cold shocked cells at 4 °C, and 2-fold in cells grown at 65 °C (Figure 5B). Surprisingly, in the FFuc strain, the *framefucA* mRNA was not affected in any of the conditions tested, suggesting that the observed increase of mRNA in the *WT* strain after cold shock was not due to transcriptional up-regulation. Moreover, in *WT* cells cold shocked at 4 °C, we observed a 2-fold increase of α-L-fucosidase activity (from 2.0 to 4.2 mU/mg at 80 and 4 °C, respectively) (Figure 5D). By contrast, the α-L-fucosidase activity remained almost constant in the cellular extract of FFuc strain.

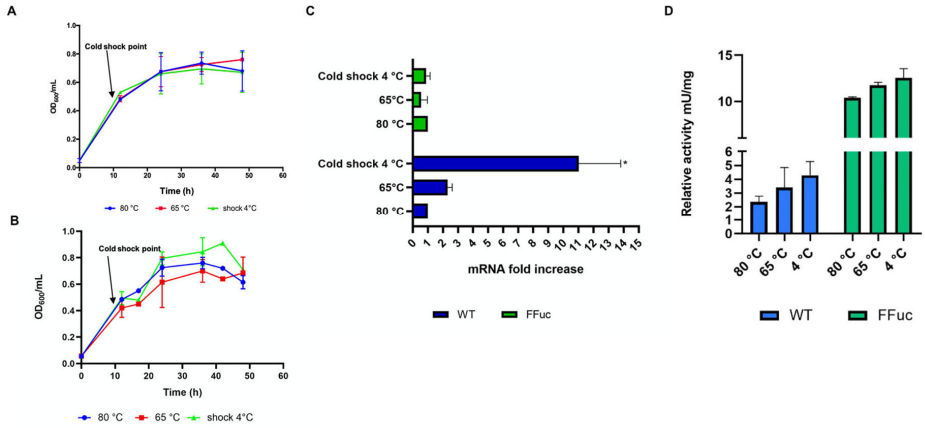


Figure 5. *S. solfataricus strains* response to cold-shock stress. (**A**) Growth curve of *S. solfataricus WT* strain at 80 °C (blue) and subjected to cold shock at 65 and 4 °C (red and green, respectively). Data are representative of three independent growth curves and are expressed as the mean ± SD. (**B**) Growth curve of *S. Solfataricus* FFuc strain at 80 °C (blue) and subjected to cold shock at 65 and 4 °C (red and green, respectively). (**C**) Real-time PCR of *fucA* and *framefucA* mRNA. The amount of mRNA in each strain grown at 80 °C was set to 1 and the other values were normalized accordingly. Data are representative of three measurements and are expressed as the mean ± SD. Statistical significance was performed using the two-tailed paired Student's *t* test (* $p < 0.05$) and ANOVA Dunnett's test ($p < 0.0013$). (**D**) Activity assays on *WT* and FFuc *S. solfataricus* cellular extracts on 4NP-α-L-fucopyranoside. Relative activity of α-L-fucosidase is expressed as mU/mg.

Following these results, we decided to evaluate the effect of UV irradiation, another stressor commonly used for *S. solfataricus* [36,37] on both strains. We analyzed the cells' viability after 60 J/m² UV doses of UV-C (254 nm) [38] by monitoring the growth curve of the *WT* and FFuc strains after UV irradiation (Figure 6). After UV irradiation, the strains

showed the same growth rate and morphological aspect of the controls (Figure 6A,B). Real-time PCR analysis revealed that the amounts in mRNA of both *fucA* and *framefucA* were reduced after UV irradiation (2- and 10-fold respectively) (Figure 6C), confirming a transcriptional down-regulation of this gene after UV irradiation as already reported [36]. The enzymatic activity assays revealed that in the FFuc strain the decrease of the transcript well correlated with the decrease of the enzymatic activity (16.5 vs. 6.8 mU/mg, control and UV, respectively). Instead, in the *WT* strain, the enzymatic activity is comparable to that of the control (3.3 and 4.0 mU/mg, control and UV respectively), despite a 2-fold decrease of the transcript level (Figure 6D).

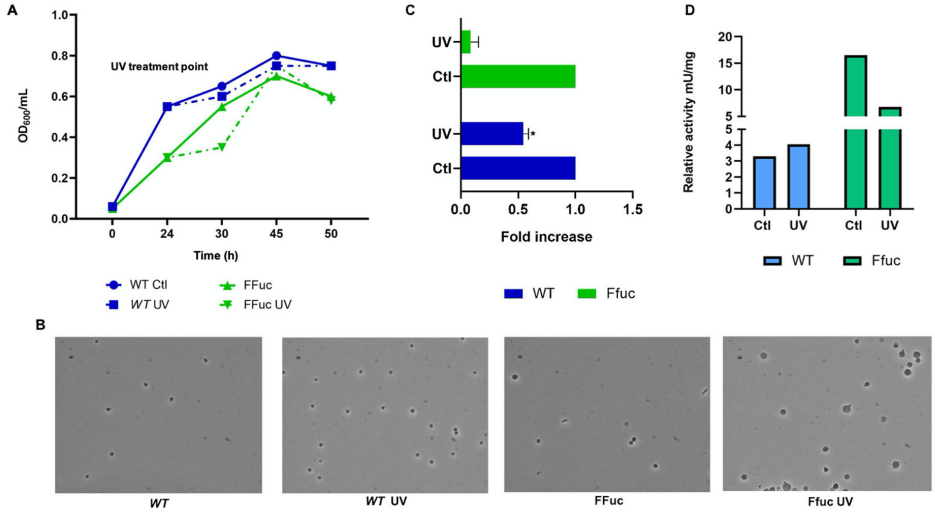


Figure 6. Response of *S. solfataricus* to UV irradiation. (**A**) Representative growth curve of *S. solfataricus* WT (blue lines) and FFuc (green lines) in standard conditions and after UV irradiation. (**B**) Contrast phase microscopy of cell cultures harvested after 50 h. (**C**) Real-time PCR of *fucA* and *framefucA* mRNA. The level of each mRNA was normalized against 16S (internal control). The mRNA amount of the non-treated *WT* and FFuc was set to 1 and the other values were normalized accordingly. Data are representative of three measurements and are expressed as the mean ± SD. Statistical significance was performed using the two-tailed paired Student's *t* test. (* $p < 0.05$). (**D**) Activity assays of WT and FFuc cellular extracts on 4NP-α-L-fucopyranoside. Relative activity of α-L-fucosidase is expressed as mU/mg.

2.3.2. Carbon Sources

S. solfataricus possesses a broad capacity for the degradation of different polymeric sugars as documented by the presence of 28 glycoside hydrolases in its genome (according to http://www.cazy.org (accessed on 4 March 2021)) [39,40]. Plant material originating from woodland areas surrounding Pisciarelli solfatara [41], mainly contribute to the presence of different polysaccharides and glycoconjugates in the organism's natural habitat. Thus, *S. solfataricus* adapted its metabolism to these environmental conditions, but relatively little is known about the function of these enzymes in vivo. However, several genes encoding for glycosyl hydrolases, namely an α-glucosidase (SSO3051), a β-glucuronidase (SSO3036), a β-xylosidase (SSO3032), and the clustered α-xylosidase (XylS) and α-glycosidase (Ssβ-gly) (SSO3022 and SSO3019, respectively), map close to *fucA* and are likely to be involved in the degradation of polysaccharides for energy metabolism [42–49]. To evaluate whether the transcription and the translation of the *fucA* gene could be affected by growing *S. solfataricus* with different carbon sources, the *WT* and FFuc strains were grown in rich and minimal media supplemented with different sugars. As shown in Figure 7A,B, both strains reached the stationary phase after 50 h in the two rich media supplemented with sucrose or fucose, respectively. By contrast, they both showed an extremely slow

growth in all three minimal media supplemented with sucrose, fucose and hydrolyzed xyloglucan. Real-time PCR analysis performed on the mRNA extracted after 90 h of growth revealed a slight transcription down-regulation for both strains grown in YCF, minimal media with sucrose or fucose (Figure 7C). By contrast, interestingly, we observed a 10-fold increase of *fucA* mRNA in *WT* cells grown in minimal medium supplemented with hydrolyzed xyloglucan (Figure 7C, blue bars). A 2-fold increase of *framefucA* mRNA has been observed also in the FFuc mutant (Figure 7C, green bars). Furthermore, in cells from *WT* we found a 2-fold increase of α-L-fucosidase enzymatic units when compared to that found in YCS (from 4.0 to 7.6 mU/mg). Instead, in FFuc strain the increase is 1.3-fold (from 16.2 to 21.0 mU/mg) (Figure 7D).

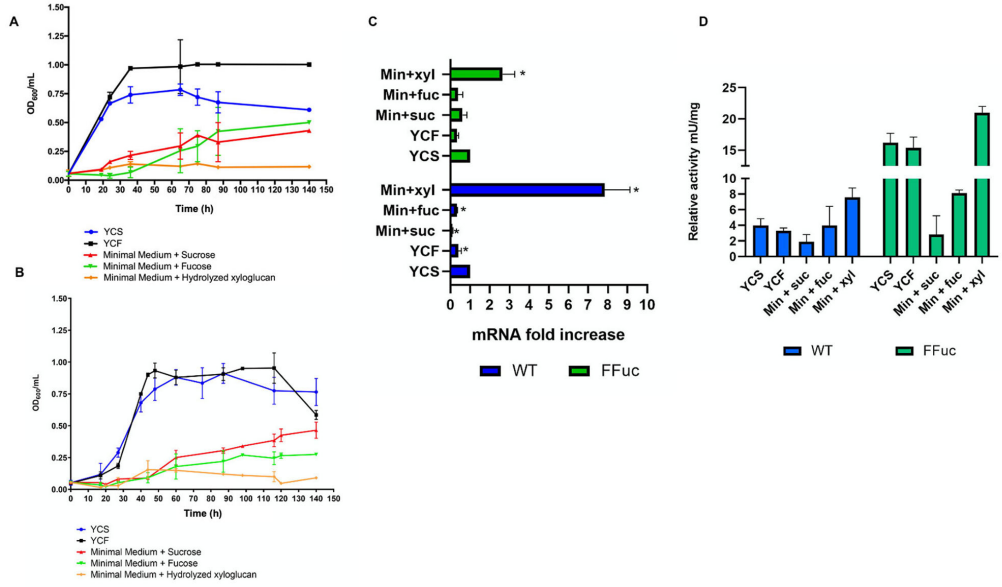


Figure 7. *S. solfataricus* strains growth with different carbon sources. (**A**) Representative growth curves of *S. solfataricus WT* strain in different media: YCS (blu), YCF (black), Minimal Medium + Sucrose (red), Minimal Medium + Fucose (green) and Minimal Medium + Hydrolyzed Xyloglucan (orange). Data are representative of three independent growth curves and are expressed as the mean ± SD. (**B**) Representative growth curves of *S. solfataricus FFuc* strain in different media: YCS (blu), YCF (black), Minimal Medium+ Sucrose (red), Minimal Medium + Fucose (green) and Minimal Medium + Hydrolyzed Xyloglucan (orange). (**C**) Real-time PCR of *fucA* and *framefucA* mRNA. The level of each mRNA was normalized against 16S (internal control). The mRNA amount in YCS was set to 1 and the other values were normalized accordingly. Data are representative of three measurements and are expressed as the mean ± SD. Statistical significance was performed using the two-tailed paired Student's *t* test (* $p < 0.05$) and ANOVA Dunnett's test ($p < 0.0001$). (**D**) Activity assays on WT and FFuc *S. solfataricus* cellular extracts on 4NP-α-L-fucopyranoside. The relative activity of α-L-fucosidase is expressed as mU/mg.

3. Discussion

Increasing evidence suggests that the flexibility of translation may increase microbial fitness under certain conditions [27]. This could be particularly relevant in extreme environments, which are subjected to sudden environmental changes, as showed in Figure 4D, to which microorganisms must rapidly adapt. PRF is one of the forms of recoding that regulates and enriches gene expression. However, the physiological significance of PRF has been assigned only to a minority of the cellular genes while for most of them it is still uncertain [1,6,13,14,50]. The reason why *fucA* expression is regulated by −1 PRF in *S. solfataricus* is not known. The polypeptide encoded by the smaller ORF SSO11867 could

never be detected by Western Blot or proteomic analyses [20,26,51] In addition, the modelling of Ssαfuc on the high-resolution crystal structure of the α-L-fucosidase from *Thermotoga maritima* showed that the N-terminal polypeptide is not an independent domain [22] and we have shown that SSO11867 includes essential catalytic residues [18], excluding the possibility that a functional α-L-fucosidase can be obtained from the C-terminal ORF SSO3060 alone. Therefore, several lines of evidence allowed us to exclude that −1 PRF is used to set the ratio of the two polypeptides, rather we suggested that this translational mechanism might be required to control the expression level of the α-L-fucosidase [20].

To get more insights into the mechanism of regulation of *fucA* in *S. solfataricus*, we compared the transcript level and the α-L-fucosidase activity of the *WT* strain with those of the FFuc strain in which *fucA*, being full-length (*framefucA*), is not under the regulation of −1 PRF.

Here, we found that the mRNA level of *framefucA* in FFuc strain is 100-fold higher than that of the *WT* in standard conditions. It is well known that faulty mRNAs with PTC are recognized and degraded by NMD [28–30]. These data suggested that the wild type mRNA is recognized as an mRNA containing a non-sense mutation and initiated towards a degradation pathway of NMD [30]. However, the presence of the full-length α-L-fucosidase [20] and of the α-L-fucosidase activity suggested that low level of −1 PRF occurred in standard conditions. In contrast, the mRNA of the FFuc strain, as expected for a full-length gene, is stable and efficiently translated as suggested by the higher α-L-fucosidase activity in cellular extracts of the FFuc strain.

Therefore, we decided to analyze the behavior of the two strains under different stress conditions. Indeed, this comparison could provide useful information on the regulation of *fucA* in vivo. Surprisingly, after cold shock, we observed a 10-fold increase of the wild type transcript in *WT* strain, while in FFuc the full-length transcript remained constant. We also observed an increase in the α-L-fucosidase activity of 2-fold. Since the putative regulatory transcriptional sequences of the wild type and mutant genes are the same, this suggests that the increase in wild type transcript was not due to transcriptional up-regulation. It has been proposed that recognition of a nonsense mRNA could depend on translation [52] and that mRNA depletion is a consequence of the appearance of long tracts of mRNA that are unprotected by scanning ribosomes [28], which, by binding to the mRNA, have a protective effect on its stability [53]. Thus, we explain the higher amounts of fucA mRNA with an increment of translating ribosomes, which stabilize fucA mRNA and lead to the increased transcript level revealed by Real-time PCR. These results are consistent with the proposed strategy adopted by the Sec codon to avoid detection as PTC by the NMD surveillance pathway, linking the selenoprotein synthesis to the efficiency of Sec incorporation. Under the conditions of adequate dietary selenium, when Sec-tRNASec levels are high, NMD presumably does not occur because Sec incorporation is efficiently out-competing translation termination [54–56]. Our data show that cold shock had an effect in vivo on the *fucA* gene, which is expressed by −1 PRF. Although both *fucA* mRNA abundance and α-L-fucosidase specific activity increase after cold shock, the variation is not comparable (10- vs. 2-fold in Figure 5C,D). Possibly, the increment rates for mRNA stabilization and protein translation is different, they depend on the −1 PRF efficiency, or α-L-fucosidase specific activity observed in these conditions was enough for the required biological effect. The involvement of α-L-fucosidase in cold shock is not clear and requires further studies that go beyond the aims of this work.

After UV irradiation, we observed a decrease in the transcript level for both wild type and mutant genes, clearly indicating that *fucA* is subjected to transcriptional down-regulation under this stress condition as already reported [36]. On the contrary, when *WT* was grown in the presence of hydrolyzed xyloglucan, we observed a 10-fold increase in the level of wild type mRNA and an increase of 2-fold in α-L-fucosidase activity. We observed also a slight (2-fold) increase in transcription and enzymatic activity (1.3-fold) in the FFuc strain. These data suggest that in these conditions transcriptional and translational up-regulation occur in both the *WT* and FFuc strains. In the full-length gene the increase in

the transcript level was much lower than in the wild type gene, suggesting that, as we proposed for the cold shock experiments, the translating ribosomes stabilized the mRNA of the wild type. The increasing demand of α-L-fucosidase in the presence of xyloglucan oligosaccharides was not surprising. Several α-L-fucosidases belonging to glycoside hydrolases family GH29 in the Carbohydrate Active enZYmes database (http://www.cazy.org/ (accessed on 4 March 2021)) [39–58]. In addition, we have previously reported that in *S. solfataricus* the α-xylosidase XylS and the β-glycosidase Ssβ-gly, hydrolyzed tamarind seed xyloglucan oligosaccharides in vitro [48]. More recently, we observed that the α-L-fucosidase is able to remove the fucose residues from fucosylated xyloglucan oligosaccharides [59], suggesting that the three enzymes cooperate for the hydrolysis of xyloglucan oligosaccharides [17] like it has been suggested in *T. maritima* [49].

We propose that we observed an increase of the wild type mRNA in grown conditions in which the α-L-fucosidase is required due to the stabilizing effect of the ribosomes which translate through −1 PRF the interrupted gene. Why the α-L-fucosidase is regulated in these, or other, grown conditions it certainly deserves further study, in order to provide new information on the possible link between the mechanism of −1 PRF in Archaea and increased fitness in extreme environments.

4. Materials and Methods

4.1. Culture Media

YCS: Brock's salt medium supplemented with yeast extract (0.1%), casamino acids (0.1%), and sucrose (0.1%) [60].

YC: Brock's salt medium supplemented with yeast extract (0.1%) and casamino acids (0.1%) [60].

YCF: Brock's salt medium supplemented with yeast extract (0.1%), casamino acids (0.1%), and fucose (0.1%) [60].

4.2. Strains and Growth Conditions

S. solfataricus strain P2 (DSM1617) and strain 98/2 (PBL2000) were considered indifferently as wild type strains [61].

S. solfataricus strain PBL2025 is a natural deletion mutant of strain 98/2, in which a fragment of about 50 kb of the chromosome containing ~50 ORFs (SSO3004-3050) is missing [62].

The mutant strains Del and FFuc were outsourced by the Gene Expression company (Gene Expression Center for Biotechnology, University of Nebraska-Lincoln, Lincoln, NE, USA). The mutants were obtained by using the PBL2025 as the parental strain. The two ORFs encoding for the α-L-fucosidase are located on a DNA region of 1487 bp. The mutant strain Del was obtained by deleting 444 bp of the *fucA* wild type gene (from position 195 to position 635), which includes the region of overlap between the two ORFs. The FFuc mutant strain was obtained by replacing the same 444 bp DNA sequence of the wild type with the DNA of the *framefucA* mutant, in which the two ORFs were carried on the same reading frame through site directed mutagenesis. The Del mutant strain has been controlled by PCR. The internal deletion of 444 bp determined the insertion of a stop codon after 71 amino acids from the ATG, preventing the translation of a full-length protein.

Unless otherwise indicated, *S. solfataricus* strains were grown at 80 °C, pH 3.5 in Brock's salt medium supplemented with yeast extract, sucrose, and casamino acids (0.1% each) [60]. The growth of cells was monitored spectrophotometrically at 600 nm and the cells were harvested at the early stationary phase (0.7–1.0 OD) by centrifugation at 5000× *g* for 15 min at 4 °C.

4.3. Growth of S. solfataricus P2 in Pisciarelli Solfatara Pool

S. solfataricus wild type, grown up to 0.17 OD_{600nm} in 100 mL of YCS in long neck flasks, was divided into 3 samples: (i) in the first sample, the culture was incubated in the

Pisciarelli solfatara by using tubes having a 0.22 μm filter at the top of the cap, which allowed the exchange of trace elements present in the solfatara water but not of microorganisms and sediments; (ii) in the second sample the cells were incubated in the same tubes but in controlled conditions (in the lab); (iii) in the third sample the cells were incubated in controlled conditions (in the lab) but in a standard laboratory 250 mL flask. The temperature of the pool was monitored for the duration of the experiment using a thermometer inserted into the device. From the time of incubation in Pisciarelli and for the following three days, the OD_{600nm} have been measured in the laboratory controls.

4.4. Growths in Different Carbon Sources

For each culture, *S. solfataricus WT* and FFuc strains were inoculated in long neck flasks in 100 mL of YCS. Growth rate was monitored spectrophotometrically at 600 nm. Each culture (initial OD_{600}: 0.04) were incubated at 80 °C, pH 3.5 and under shaking at 160 rpm overnight (ON). Once the culture reached the early exponential phase (0.4–0.5 OD_{600}), it was diluted to a value of 0.05 OD_{600} in 250 mL of fresh media and growth at 80 °C.

1. Rich medium (YCS): Brock's salt medium supplemented with yeast extract (0.1%), casamino acids (0.1%), plus sucrose (0.1%).

2. Minimal medium + Sucrose: Brock's salt medium supplemented with sucrose (0.1%).

3. Minimal medium + Fucose: Brock's salt medium supplemented with fucose (0.1%).

4. Minimal medium + Hydrolyzed Xyloglucan: Brock's salt medium supplemented with non fucosylated hydrolyzed Xyloglucan mix (0.1%).

5. Rich medium with fucose (YCF): Brock's salt medium supplemented with yeast extract, casamino acids and fucose (0.1% each).

Growth rate was monitored spectrophotometrically at 600 nm and cultures were harvested after 90 h (late exponential phase), centrifuged at 3500× *g* for 5 min and stored at −20 °C until use.

4.5. Cold Shock

S. solfataricus WT and FFuc strains were inoculated in long neck flasks in 100 mL of YCS and incubated at 80 °C, pH 3.5 and under shaking at 160 rpm overnight (ON). Growth rate was monitored spectrophotometrically at 600 nm. Once the culture reached the logarithmic phase of growth (0.4–0.5 OD_{600}), it was diluted to a value of 0.05 OD_{600} in three aliquots of 250 mL of fresh YCS medium and cultured at 80 °C up to 0.5 OD_{600}; once arrived in early exponential phase, each strain culture was incubated (i) at 65 °C (ii) at 4 °C for 2 h and then again at 80 °C. As control one culture of each strain was grown at 80 °C as described above. For each culture, growth rate was monitored spectrophotometrically using a Cary 100 (Agilent, Santa Clara, CA, USA) at 600 nm and cells were harvested during the early stationary phase, at 0.8 OD_{600}. Cells were centrifuged at 3500× *g* for 15 min and pellets stored at −20 °C.

4.6. UV Irradiation

S. solfataricus WT and FFuc strains were grown in 100 mL of YCS up to 0.4–0.5 OD_{600}. An aliquot of 50 mL was put in ice and used for UV-irradiation. Irradiation was performed by means of a UV-C lamp (model EF-280C/FE 230-volt 50 Hz 0.34 AMPS. Spectroline, Westbury, NY, USA) set at λ = 254 nm with a power of about 6.00 W/m^2. The UV flux at the surface of cells was measured with a UV digital radiometer (HD 2102.2. Delta OHM, Caselle di Selvazzano (PD), Italy). The experiment was carried out in sterile conditions: in particular, 5 aliquots of 10 mL were transferred to a plastic petri dish (110 mm) and irradiated with UV light for 20 s at 245 nm (corresponding to an absorbed dose of 60 J/m^2) while shaking the culture carefully. The treated cultures were stored in the dark on ice for

30 min and then incubated at 80 °C under shaking. Cell growth was spectrophotometrically monitored, and the samples were harvested after 50 h. Cells were centrifuged at 3500× *g* for 15 min and pellets stored at −20 °C until use.

4.7. RNA Extraction and Real-Time PCR

Total cellular RNA was extracted from *S. solfataricus WT* and FFuc cells, according to the Qiagen's RNeasy mini protocol, with a DNase step treatment, at 37 °C for 20 min, when necessary. Total RNA concentrations were estimated with Qubit 4 Fluorometer using the Qubit™ RNA HS Assay Kit (Thermo Fisher, Waltham, MA, USA) and RNA quality was independently assessed by visualization on a 1.5% agarose (wt/vol) gels. cDNA was synthesized from 300–600 ng of total RNA from each sample using SuperScript™ IV VILO™ Master Mix reverse transcriptase (Thermo Fisher, Waltham, MA, USA) in a 25 μL reaction. The reverse transcription reactions were performed according to Thermo Fisher protocol. In particular, for each a reaction mixture containing 4 μL of RT reaction mix and 300–600 ng of RNA was prepared as a control the same mixture without reverse transcriptase was used. The program used was as follows:10 min at 25 °C, 10 min at 50 °C, 5 min at 85 °C.

The expression pattern of α-L-fucosidase (*fucA* and *framefucA*) for *WT* and FFuc strains, respectively, was analyzed by using gene-specific primers FucFWD: 5′-TGCCAGATAGACCAGAACAC-3′ and FucREV5′-GCCCTATGATACGAAATGCC-3′ designed to amplify a 277 bp region of the α-L-fucosidase gene (SSO11867-SSO3060). Oligos were designed using the qPCR Probes Design Tool-GenScript (GenScript.com, Piscataway, NJ, USA). Real-time polymerase chain reactions (real-time PCR) were performed by using SYBR™ Green PCR Master Mix (Thermo Fisher, Waltham, MA, USA); each reaction (20 μL) mix contained 10 μL of Mastermix, 150 nM (final concentration) of each primer and 4 μL of cDNA (10 ng). PCR products were detected using Applied Biosystems 7300/7500/7500 Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) The program used was as follows: 10 min at 95 °C, 15 s at 95 °C and 1 min at 60 °C for 40 cycles. All samples were analyzed in triplicate. *fucA* and *framefucA* gene expression profiles were normalized against 16S transcript to correct for differences in the starting amount of RNA and in the efficiency of the reverse transcription reaction. Statistical significance was performed for all qRTPCR experiments using the two-tailed paired Student's *t* test, and, for cold shock and different carbon sources growth experiments, also with ANOVA Dunnett's test. Melting curve analyses of each PCR reaction were performed to assess specificity.

4.8. Optical Microscopy

Cell aliquots were collected during exponential phase and seeded on a slide glass and analyzed with an Olympus BX51 contrast phase microscope (Olympus corporation, Shinjuku, Tokyo, Japan), using a 100× lens.

4.9. Cell Lysates Preparation

The cell pellets obtained after centrifugation were resuspended in lysis butter (20 mM potassium phosphate buffer pH 7.2, 150 mM NaCl, 0.1% Triton X 100) (1:5 *v*/*v*) and cell lysis was performed with 5 cycles of freeze and thaw. The supernatant was clarified by centrifugation at 12,000× *g* for 30 min, soluble fraction was separated, and protein concentration was measured with Bradford Protein Assay Kit (Bio-Rad, Hercules, CA, USA). Samples were kept frozen until use.

4.10. Western Blot Analyses

S. solfataricus cells were lysed as described above and lysates were separated in 8% SDS-polyacrylamide gel and transferred to PVDF membranes (Merck-Millipore, Burlington, MA, USA). Filters were blocked for 3 h at RT in 5% (*w*/*v*) non-fat milk in PBS (Phosphate-buffered saline) 1X Tween-20 0.1% (TPBS).

For α-L-fucosidase, the filter was incubated with anti-FucA primary antibody (1:500, PRIMM srl., Milan, Italy). After several washings in TPBS, membranes were incubated with secondary antibody against mouse (1:50,000 PIERCE. Thermo Fisher, Waltham, MA, USA) linked to horseradish peroxidase, and signals were visualized by chemiluminescence (ECL, Amersham, Little Chalfont, UK).

For fucosylated glycoproteins the filter was incubated with the UEAI-HRP labelled antibody (0.1 mg/mL) for 2 h. After several washings in TPBS signals were visualized by chemiluminescence (ECL, Amersham, Little Chalfont, UK).

4.11. Alpha-L-Fucosidase Activity Assay

α-L-fucosidase activity assay was performed at 75 °C in 50 mM sodium phosphate buffer at pH 6.5, using 4-nitrophenyl-α-L-fucopyranoside (4NP-α-L-Fuc) substrate at the final concentration of 4 mM and different amount of *S. solfataricus WT* and FFuc cellular extracts. The reaction was blocked by adding 0.5 M Na_2CO_3 and the product, 4-nitrophenolate (4NP), was detected spectrophotometrically at 405 nm. The extinction coefficient used was = 18.2 $mM^{-1}cm^{-1}$. For all assays, spontaneous hydrolysis of the substrate was subtracted by using appropriate blank mixtures without cell lysates. Enzymatic activity assays have been performed in duplicate and reported as the mean ± SD. One unit (U) of enzymatic activity was defined as the amount of enzyme that released 1 umoL of 4NP per min at the conditions described. The units of enzymatic activity have been normalized for mg of total proteins in the cellular extracts.

5. Conclusions

Noticeably, α-L-fucosidases are extremely rare in Archaea and, up to now, *fucA* is the only gene known so far that is expressed by −1 PRF in this Domain of life. Several functional interrupted genes were identified in *S. solfataricus* [26] suggesting that in Archaea, more genes could be regulated by translational recoding, such as those in viral genomes or encoding for proteins with no enzymatic activity, but that have not yet been identified, possibly because of the difficulty of isolating and characterizing them. Under this point of view, the α-L-fucosidase activity is a useful molecular tool to study −1 PRF as it be easily assayed in vitro [17]. Here, the analysis of different growth conditions showed that cold shock and the presence of xyloglucan oligosaccharides increased up to 10-fold the mRNA abundance of *fucA*, while the full-length control gene showed mRNA levels similar and much less increased, respectively, if compared to standard growth conditions. We propose that the reason of the more abundant mRNA is due to the presence of ribosomes performing −1 PRF and thereby preventing its degradation. This may suggest that cold shock and xyloglucan oligosaccharides induce −1 PRF, but further studies, going beyond the aims of this work, are required.

It has been already postulated that the flexibility of the genetic code decoding is a trait selected during evolution to benefit microorganisms under certain conditions [27]. It is tempting to speculate that this regulation at translational level might be advantageous in extreme environments, which are often spots (e.g., for hydrothermal vents, solfataras, acidic/basic/salty ponds, etc.) located in places dominated by mild conditions. In these "extreme" sites, microbial communities might encounter sudden and reversible changes of the optimal growth conditions more frequently than microbes living in stable conditions. Therefore, translational recoding could be a way to maintain the expression of certain genes latent, and up- or down-regulate them under specific conditions. Approaches

of system biology on the large amounts of available (meta) genomic data from extremophiles might open new avenues to the study of translational recoding in this domain of life [63].

Author Contributions: Conceptualization, B.C.-P. and M.M.; methodology, F.D.L., R.I., A.S., R.G., S.C., P.D.D.; investigation, F.D.L., R.I., A.S., R.G., N.C., L.M., R.A., A.C., A.T., F.L.; writing—original draft preparation, F.D.L., B.C.-P. and M.M.; funding acquisition, B.C.-P. and M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Italian Space Agency: "Life In Space (OPPS)" project (ASI N. 2019-3-U.O.) and "Esobiologia e ambienti estremi: dalla Chimica delle Molecole alla Biologia degli Estremofili (ECMB)" project (ASI N. 2014-026-R.O.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Acknowledgments: We are grateful to Mauro Di Fenza (IBBR-CNR) for the critical reading of the manuscript, and to Giovanni del Monaco, Chiara Nobile and Marco Petruzziello (IBBR-CNR) for administrative and technical assistance.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Samples are not available from the authors.

References

1. Atkins, J.F.; Loughran, G.; Bhatt, P.R.; Firth, A.E.; Baranov, P.V. Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* **2016**, *44*, 7007–7078, doi:10.1093/nar/gkw530.
2. Bekaert, M.; Firth, A.E.; Zhang, Y.; Gladyshev, V.N.; Atkins, J.F.; Baranov, P.V. Recode-2: New design, new search tools, and many more genes. *Nucleic Acids Res.* **2009**, *38*, D69–D74, doi:10.1093/nar/gkp788.
3. Gesteland, R.F.; Atkins, J.F. Recoding: Dynamic reprogramming of translation. *Annu. Rev. Biochem.* **1996**, *65*, 741–768.
4. Farabaugh, P.J. Programmed Translational Frameshifting. *Annu. Rev. Genet.* **1996**, *30*, 507–528, doi:10.1146/annurev.genet.30.1.507.
5. Namy, O.; Rousset, J.-P.; Napthine, S.; Brierley, I. Reprogrammed Genetic Decoding in Cellular Gene Expression. *Mol. Cell* **2004**, *13*, 157–168, doi:10.1016/s1097-2765(04)00031-0.
6. Rodnina, M.V.; Korniy, N.; Klimova, M.; Karki, P.; Peng, B.-Z.; Senyushkina, T.; Belardinelli, R.; Maracci, C.; Wohlgemuth, I.; Samatova, E.; et al. Translational recoding: Canonical translation mechanisms reinterpreted. *Nucleic Acids Res.* **2020**, *48*, 1056–1067, doi:10.1093/nar/gkz783.
7. Bertram, G.; Innes, S.; Minella, O.; Richardson, J.P.; Stansfield, I. Endless possibilities: Translation termination and stop codon recognition. *Microbiology* **2001**, *147*, 255–269, doi:10.1099/00221287-147-2-255.
8. Nicholas, P.R.; Jean-François Brugère, J.F.; Atkins, P.W.; O'Toole, G.B. Pyrrolysine in archaea: A 22nd amino acid encoded through a genetic code expansion. *Emerg. Top. Life Sci.* **2018**, 607–618.
9. Ambrogelly, A.; Palioura, S.; Söll, D. Natural expansion of the genetic code. *Nat. Chem. Biol.* **2006**, *3*, 29–35, doi:10.1038/nchembio847.
10. Eswarappa, S.M.; Potdar, A.A.; Koch, W.J.; Fan, Y.; Vasu, K.; Lindner, D.; Willard, B.; Graham, L.M.; DiCorleto, P.E.; Fox, P.L. Programmed Translational Readthrough Generates Antiangiogenic VEGF-Ax. *Cell* **2014**, *157*, 1605–1618, doi:10.1016/j.cell.2014.04.033.
11. Keeling, K.M.; Xue, X.; Gunn, G.; Bedwell, D.M. Therapeutics Based on Stop Codon Readthrough. *Annu. Rev. Genom. Hum. Genet.* **2014**, *15*, 371–394, doi:10.1146/annurev-genom-091212-153527.
12. Plant, E.P.; Rakauskaitė, R.; Taylor, D.R.; Dinman, J.D. Achieving a Golden Mean: Mechanisms by Which Coronaviruses Ensure Synthesis of the Correct Stoichiometric Ratios of Viral Proteins. *J. Virol.* **2010**, *84*, 4330–4340, doi:10.1128/jvi.02480-09.
13. Kelly, J.A.; Olson, A.N.; Neupane, K.; Munshi, S.; San Emeterio, J.; Pollack, L.; Woodside, M.T.; Dinman, J.D. Structural and functional conservation of the programmed −1 ribosomal frameshift signal of SARS coronavirus 2 (SARS-CoV-2) *J. Biol. Chem.* **2020**, *295,* 10741.
14. Kurian, L.; Palanimurugan, R.; Gödderz, D.; Dohmen, R.J. Polyamine sensing by nascent ornithine decarboxylase antizyme stimulates decoding of its mRNA. *Nat. Cell Biol.* **2011**, *477*, 490–494, doi:10.1038/nature10393.
15. Rother, M.; Quitzke, V. Selenoprotein synthesis and regulation in Archaea. *Biochim. Biophys. Acta (BBA) Gen. Subj.* **2018**, *1862*, 2451–2462, doi:10.1016/j.bbagen.2018.04.008.
16. Cobucci-Ponzano, B.; Conte, F.; Rossi, M.; Moracci, M. The α-L-fucosidase from Sulfolobus solfataricus. *Extremophiles* **2007**, *12*, 61–68, doi:10.1007/s00792-007-0105-y.

17. Cobucci-Ponzano, B.; Trincone, A.; Giordano, A.; Rossi, M.; Moracci, M. Identification of an archaeal α-L-fucosidase en-coded by an interrupted gene: Production of a functional enzyme by mutations mimicking programmed −1 frameshifting *J. Biol. Chem.* **2003**, *278*, 14622–14631.
18. Cobucci-Ponzano, B.; Trincone, A.; Giordano, A.; Rossi, M.; Moracci, M. Identification of the catalytic nucleophile of the family 29 α-L-fucosidase from Sulfolobussolfataricus via chemical rescue of an inactive mutant *Biochemistry* **2003**, *42*, 9525–9531.
19. Cobucci-Ponzano, B.; Mazzone, M.; Rossi, M.; Moracci, M. Probing the Catalytically Essential Residues of the α-L-Fucosidase from the Hyperthermophilic ArchaeonSulfolobus solfataricus. *Biochemestry* **2005**, *44*, 6331–6342, doi:10.1021/bi047495f.
20. Cobucci-Ponzano, B.; Conte, F.; Benelli, D.; Londei, P.; Flagiello, A.; Monti, M.; Pucci, P.; Rossi, M.; Moracci, M. The gene of an archaeal α-L-fucosidase is expressed by translational frameshifting *Nucleic Acids Res.* **2006**, *34*, 4258–4268.
21. Cobucci-Ponzano, B.; Rossi, M.; Moracci, M. Recoding in Archaea. *Mol. Microbiol.* **2004**, *55*, 339–348, doi:10.1111/j.1365-2958.2004.04400.x.
22. Rosano, C.; Zuccotti, S.; Cobucci-Ponzano, B.; Mazzone, M.; Rossi, M.; Moracci, M.; Petoukhov, M.V.; Svergun, D.I.; Bolognesi, M. Structural characterization of the nonameric assembly of an Archaeal alpha-L-fucosidase by synchrotron small angle X-rays cattering. *Biochem. Biophys. Res. Commun*. **2004**, *320*, 176–182.
23. Pietilä, M.K.; Laurinmäki, P.; Russell, D.A.; Ko, C.-C.; Jacobs-Sera, D.; Butcher, S.J.; Bamford, D.H.; Hendrix, R.W. Insights into Head-Tailed Viruses Infecting Extremely Halophilic Archaea. *J. Virol.* **2013**, *87*, 3248–3260, doi:10.1128/jvi.03397-12.
24. Senčilo, A.; Jacobs-Sera, D.; Russell, D.A.; Ko, C.-C.; Bowman, C.A.; Atanasova, N.S.; Österlund, E.; Oksanen, H.M.; Bamford, D.H.; Hatfull, G.F.; et al. Snapshot of haloarchaeal tailed virus genomes. *RNA Biol.* **2013**, *10*, 803–816, doi:10.4161/rna.24045.
25. Antonov, I.; Coakley, A.; Atkins, J.F.; Baranov, P.V.; Borodovsky, M. Identification of the nature of reading frame transitions observed in prokaryotic genomes. *Nucleic Acids Res.* **2013**, *41*, 6514–6530, doi:10.1093/nar/gkt274.
26. Cobucci-Ponzano, B.; Guzzini, L.; Benelli, D.; Londei, P.; Perrodou, E.; Lecompte, O.; Tran, D.; Sun, J.; Wei, J.; Mathur, E.J.; et al. Functional Characterization and High-Throughput Proteomic Analysis of Interrupted Genes in the ArchaeonSulfolobus solfataricus. *J. Proteome Res.* **2010**, *9*, 2496–2507, doi:10.1021/pr901166q.
27. Ling, J.; O'Donoghue, P.; Söll, D. Genetic code flexibility in microorganisms: Novel mechanisms and impact on physiology. *Nat. Rev. Genet*. **2015**, *13*, 707–721, doi:10.1038/nrmicro3568.
28. Brogna, S.; McLeod, T.; Petric, M. The Meaning of NMD: Translate or Perish. *Trends Genet.* **2016**, *32*, 395–407, doi:10.1016/j.tig.2016.04.007.
29. Celik, A.; He, F.; Jacobson, A. NMD monitors translational fidelity 24/7. *Curr. Genet.* **2017**, *63*, 1007–1010.
30. Clouet-D'Orval, B.; Batista, M.; Bouvier, M.; Quentin, Y.; Fichant, G.; Marchfelder, A.; Maier, L.-K. Insights into RNA-processing pathways and associated RNA-degrading enzymes in Archaea. *FEMS Microbiol. Rev.* **2018**, *42*, 579–613, doi:10.1093/femsre/fuy016.
31. De Rosa, M.; Gambacorta, A.; Bu'Lock, J. D. Extremely thermophilic acidophilic bacteria convergent with Sulfolobus aci-docaldarius. *J. Gen. Microbiol.* **1975**, *86*, I56–164.
32. Van Wolferen, M.; Wagner, A.; van der Does, C.; Albers, S.V. The archaeal Ced system imports DNA. *Proc. Natl. Acad. Sci. USA* **2016**, *13*, 2496–2501.
33. Fröls, S.; Ajon, M.; Wagner, M.; Teichmann, D.; Zolghadr, B.; Folea, M.; Boekema, E.J.; Driessen, A.J.M.; Schleper, C.; Albers, S.V. UV inducible cellular aggregation of the hyperthermophilic archaeon Sulfolobus solfataricus is mediated by pili for-mation. *Mol. Microbiol.* **2008**, *70*, 938–952.
34. Barak, Z.; Gallant, J.; Lindsley, D.; Kwieciszewki, B.; Heidel, D. Enhanced Ribosome Frameshifting in Stationary Phase Cells. *J. Mol. Biol.* **1996**, *263*, 140–148, doi:10.1006/jmbi.1996.0565.
35. Wenthzel, A.-M.K.; Stancek, M.; Isaksson, L.A. Growth phase dependent stop codon readthrough and shift of translation reading frame inEscherichia coli. *FEBS Lett.* **1998**, *421*, 237–242, doi:10.1016/s0014-5793(97)01570-6.
36. Götz, D.; Paytubi, S.; Munro, S.; Lundgren, M.; Bernander, R.; White, M.F. Responses of hyperthermophilic crenarchaea to UV irradiation. *Genome Biol.* **2007**, *8*, R220–18, doi:10.1186/gb-2007-8-10-r220.
37. Rolfsmeier, M.L.; Laughery, M.F.; Haseltine, C.A. Repair of DNA Double-Strand Breaks following UV Damage in Three Sulfolobussolfataricus Strains. *J. Bacteriol.* **2010**, *192*, 4954–4962, doi:10.1128/jb.00667-10.
38. Di Donato, P.; Romano, I.; Mastascusa, V.; Poli, A.; Orlando, P.; Pugliese, M.; Nicolaus, B. Survival and Adaptation of the Thermophilic Species Geobacillus thermantarcticus in Simulated Spatial Conditions. *Orig. Life Evol. Biosph.* **2017**, *48*, 141–158, doi:10.1007/s11084-017-9540-7.
39. Lombard, V.; Ramulu, H.G.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490–D495, doi:10.1093/nar/gkt1178.
40. Strazzulli, A.; Cobucci-Ponzano, B.; Iacono, R.; Giglio, R.; Maurelli, L.; Curci, N.; Schiano-di-Cola, C.; Santangelo, A.; Contursi, P.; Lombard, V.; et al. Discovery of hyperstable carbohy-drate-active enzymes through metagenomics of extreme environments. *FEBS J.* **2020**, *287*, 1116–1137.
41. Brock, T.D. Thermophilic Microorganisms and Life at High Temperatures. *Ascomycete Syst.* **1978**, 465, doi:10.1007/978-1-4612-6284-8.
42. Cobucci-Ponzano, B.; Moracci, M.; Di Lauro, B.; Ciaramella, M.; D'Avino, R.; Rossi, M. Ionic network at the C-terminus of the beta-glycosidase from the hyperthermophilic archaeon Sulfolobus solfataricus: Functional role in the quaternary structure thermal stabilization. *Proteins Struct. Funct. Genet.* **2002**, *48*, 98–106.

43. Iacono, R.; Strazzulli, A.; Maurelli, L.; Curci, N.; Casillo, A.; Corsaro, M.M.; Moracci, M.; Cobucci-Ponzano, B. GlcNAc De-N-Acetylase from the hyperthermophilic archaeon Sulfolobus solfataricus *Appl. Environ. Microbiol.* **2019**, *85,* 01879.
44. Cobucci-Ponzano, B.; Strazzulli, A.; Iacono, R.; Masturzo, G.; Giglio, R.; Rossi, M.; Moracci, M. Novel thermophilic hem-icellulases for the conversion of lignocellulose for second generation biorefineries. *Enzym. Microb. Technol.* **2015**, *78*, 63–73.
45. Ferrara, M.C.; Cobucci-Ponzano, B.; Carpentieri, A.; Henrissat, B.; Rossi, M.; Amoresano, A.; Moracci, M. 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 (BBA) Gen. Subj.* **2014**, *1840*, 367–377, doi:10.1016/j.bbagen.2013.09.022.
46. Cobucci-Ponzano, B.; Conte, F.; Strazzulli, A.; Capasso, C.; Fiume, I.; Pocsfalvi, G.; Rossi, M.; Moracci, M. The molecular characterization of a novel GH38 α-mannosidase from the crenarchaeon Sulfolobus solfataricus revealed its ability in de-mannosylating glycoproteins. *Biochimestry* **2010**, *92*, 1895–1907, doi:10.1016/j.biochi.2010.07.016.
47. Cobucci-Ponzano, B.; Aurilia, V.; Riccio, G.; Henrissat, B.; Coutinho, P.M.; Strazzulli, A.; Padula, A.; Corsaro, M.M.; Pieretti, G.; Pocsfalvi, G.; et al. A new archaeal β-glycosidase from Sulfolobus solfataricus: Seeding a novel retaining β-glycan- specific glycoside hydrolase family along with the human non-lysosomal gluco-sylceramidase GBA. *J. Biol. Chem.* **2010**, *285*, 20691–20703.
48. Ausili, A.; Cobucci-Ponzano, B.; Di Lauro, B.; D'Avino, R.; Perugino, G.; Bertoli, E.; Scirè, A.; Rossi, M.; Tanfani, F.; Moracci, M. A comparative infrared spectroscopic study of glycoside hydrolases from extremophilic archaea revealed different molecular mechanisms of adaptation to high temperatures. *Proteins Struct. Funct. Bioinform.* **2007**, *67*, 991–1001, doi:10.1002/prot.21368.
49. Moracci, M.; Ponzano, B.C.; Trincone, A.; Fusco, S.; De Rosa, M.; van der Oost, J.; Sensen, C.W.; Charlebois, R.L.; Rossi, M. Identification and Molecular Characterization of the First α-Xylosidase from an Archaeon. *J. Biol. Chem.* **2000**, *275*, 22082–22089, doi:10.1074/jbc.m910392199.
50. Belew, A.T.; Meskauskas, A.; Musalgaonkar, S.; Advani, V.M.; Sulima, S.O.; Kasprzak, W.K.; Shapiro, B.A.; Dinman, J.D. Ribosomal frameshifting in the CCR5 mRNA is regulated by miRNAs and the NMD pathway. *Nat. Cell Biol.* **2014**, *512*, 265–269, doi:10.1038/nature13429.
51. Chong, P.K.; Wright, P.C. Identification and characterization of the Sulfolobus solfataricus P2 proteome. *J. Proteome Res.* **2005**, *4*, 1789–1798.
52. Raimondeau, E.; Bufton, J.C.; Schaffitzel, C. New insights into the interplay between the translation machinery and nonsense-mediated mRNA decay factors. *Biochem. Soc. Trans.* **2018**, *46*, 503–512, doi:10.1042/bst20170427.
53. Vargas-Blanco, D.A.; Shell, S.S. Regulation of mRNA Stability During Bacterial Stress Responses. *Front. Microbiol.* **2020**, *11*, 2111, doi:10.3389/fmicb.2020.02111.
54. Fletcher, J.E.; Copeland, P.R.; Driscoll, N.M. Polysome distribution of phospholipid hydroperoxide glutathione peroxidase mRNA: Evidence for a block in elongation at the UGA/selenocysteine codon. *RNA* **2000**, *6*, 1573–1584, doi:10.1017/s1355838200000625.
55. Martin, G.W.; Berry, M.J. Selenocysteine codons decrease polysome association on endogenous selenoprotein mRNAs. *Genes Cells* **2001**, *6,* 121–29.
56. Shetty, S.P.; Copeland, P.R. Selenocysteine incorporation: A trump card in the game of mRNA decay. *Biochimestry* **2015**, *114*, 97–101, doi:10.1016/j.biochi.2015.01.007.
57. Rubianes, D.; Valdivia, E.R.; Revilla, G.; Zarra, I.; Sampedro, J. Xyloglucan exoglycosidases in the monocot model Brachypodium distachyon and the conservation of xyloglucan disassembly in angiosperms. *Plant Mol. Biol.* **2019**, *100*, 495–509.
58. Larsbrink, J.; Thompson, A.J.; Lundqvist, M.; Gardner, J.G.; Davies, G.J.; Brumer, H. A complex gene locus enables xyloglucan utilization in the model saprophyteCellvibrio japonicus. *Mol. Microbiol.* **2014**, *94*, 418–433, doi:10.1111/mmi.12776.
59. Curci, N.; Strazzulli, A.; Iacono, R.; De Lise, F.; Maurelli, L.; Di Fenza, M.; Cobucci-Ponzano, B.; Moracci, M. Xyloglucan Oligosaccharides Hydrolysis by Exo-Acting Glycoside Hydrolases from Hyperthermophilic Microorganism *Saccharolobus solfataricus*. *Int. J. Mol. Sci.* **2021**, *22*, 3325.
60. Brock, T.D.; Brock, K.M.; Belly, R.T.; Weiss, R.L. Sulfolobus: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Microbiol.* **1972**, *84*, 54–68, doi:10.1007/bf00408082.
61. McCarthy, S.; Gradnigo, J.; Johnson, T.; Payne, S.; Lipzen, A.; Martin, J.; Schackwitz, W.; Moriyama, E.; Blum, P. Complete Genome Sequence of Sulfolobus solfataricus Strain 98/2 and Evolved Derivatives. *Genome Announc.* **2015**, *28*, e00549-15.
62. Koerdt, A.; Jachlewski, S.; Ghosh, A.; Wingender, J.; Siebers, B.; Albers, S.-V. Complementation of Sulfolobus solfataricus PBL2025 with an α-mannosidase: Effects on surface attachment and biofilm formation. *Extremophiles* **2011**, *16*, 115–125, doi:10.1007/s00792-011-0411-2.
63. Onofri, S.; Balucani, N.; Barone, V.; Benedetti, P.; Billi, D.; Balbi, A.; Brucato, J.R.; Cobucci-Ponzano, B.; Costanzo, G.; Rocca, N.; et al. OPPS Project Tea the Italian National Project of Astrobiology-Life in Space-Origin, Presence, Persistence of Life in Space, from Molecules to Extremophiles. *Astrobiology* **2020**, *20*, 580–582.