Enzyme promiscuity at the origin of metallo-β-lactamases and within the α/β-hydrolase superfamily

Dissertation

with the aim of achieving the degree of *Doctor rerum naturalium* (Dr. rer. nat.)

at the Faculty of Mathematics, Informatics and Natural Sciences,

Department of Biology

of Universität Hamburg.

Submitted by

Pablo Pérez-García

born October 21, 1992 in Cortes de Pallás (Valencia, Spain).

Hamburg 2020

Genehmigt vom Fachbereich Biologie der Universität Hamburg

auf Antrag von Prof. Dr. W. R. Streit.

Weiterer Gutachter der Dissertation: Prof. Dr. Bernward Bisping.

Tag der Disputation: 26.05.2020.

Abstract

Enzymes at the origin of evolution are believed to be promiscuous “all-rounders”, taking part in various primitive metabolic pathways back when organisms had relatively simple and small genomes and limited enzymatic resources. While protein evolution is associated with the specialization of activities, ancestral traces remain in the genome of some organisms. Promiscuous enzymes are of importance in protein and metabolic evolution, environmental and structural biology as well as in biotechnology.

In this work, promiscuity traits were studied for two of the most diverse and evolutionary successful protein (super-) families known: the metallo-β-lactamases (MβLs) and the α/β-hydrolases (ABHs).

The MβL Igni18 originates from the crenarchaeon *Ignicoccus hospitalis*, which has one of the smallest genomes known for a free-living organism. It was chosen as a model to elucidate structural evolution of the MβL-fold. The enzyme is extremely thermostable and highly promiscuous, thus adapted to the environmental conditions of its host´s habitat. Within this thesis, I was able to produce this crenarchaeal and ancestral protein in the methylotrophic yeast *Pichia pastoris*. The protein was crystallized, and its structure could be determined. The rather compact crystal structure of the monomer turned out to be very similar to the core of a set of specialized enzymes and therefore, Igni18 exemplifies the structure of a common ancestor. Based on these analyses, ten Protein Variable Regions (PVRs) were defined that distinguish the original, “multitasking” MβL from its modern relatives.

In the second part of this work, enzyme promiscuity in the ABH superfamily was studied within a set of metagenome-derived clones with esterase activity. Their reaction profile against a collection of 96 esters allowed the classification in three categories according to their level of promiscuity. 80 new, putative hydrolase-coding genes were described and initial characterization of 45 of them was performed.

Results from this thesis highlight the need to rethink the way metabolic pathways are being predicted in the presence of enzyme promiscuity.

Zusammenfassung

Es wird angenommen, dass Enzyme am Ursprung der Evolution promiskuitive „Alleskönner“ waren. Sie waren an verschiedenen primitiven Stoffwechselwegen beteiligt, als Organismen noch relativ einfache und kleine Genome und begrenzte enzymatische Ressourcen besaßen. Während die Proteinentwicklung mit der Spezialisierung von Aktivitäten verbunden ist, verbleiben Ahnenspuren im Genom einiger Organismen. Promiskuitive Enzyme sind für die Protein- und Stoffwechselentwicklung, die Umwelt- und Strukturbiologie sowie für die Biotechnologie von Bedeutung.

In dieser Arbeit wurden Promiskuitätsmerkmale für zwei der vielfältigsten und evolutionär erfolgreichsten bekannten Protein- (super-) Familien untersucht: die Metallo-β-Lactamasen (MβLs) und die α/β-Hydrolasen (ABHs).

Die MβL Igni18 stammt aus dem Crenarchaeon *Ignicoccus hospitalis*, das eines der kleinsten Genome aufweist, die für einen frei lebenden Organismus bekannt sind. Es wurde als Modell ausgewählt, um die strukturelle Entwicklung der MβL-Faltung aufzuklären. Das Enzym ist extrem thermostabil und sehr promiskuitiv und daher an die Umweltbedingungen des Lebensraums seines Wirts angepasst. Im Rahmen dieser Arbeit konnte ich dieses crenarchaeelle Vorgängerprotein in der methylotrophen Hefe *Pichia pastoris* herstellen. Das Protein wurde kristallisiert und seine Struktur konnte bestimmt werden. Die ziemlich kompakte Kristallstruktur des Monomers erwies sich dem Kern einer Reihe spezialisierter Enzyme als sehr ähnlich, und daher veranschaulicht Igni18 die Struktur eines gemeinsamen Vorfahren. Basierend auf diesen Analysen wurden zehn „Protein Variable Regions“ (PVRs) definiert, die das ursprüngliche „Multitasking“-MβL von seinen modernen Verwandten unterscheiden.

Im zweiten Teil dieser Arbeit wurde die Promiskuität von Enzymen in der ABH-Superfamilie in einer Reihe von Metagenomklonen mit Esteraseaktivität untersucht. Ihr Reaktionsprofil wurde gegen eine Sammlung von 96 Estern erstellt. Dies ermöglichte die Einteilung in drei Kategorien nach ihrem Promiskuitätsgrad. 80 neue vermutlich Hydrolase-kodierende Gene wurden beschrieben und 45 von ihnen wurden zunächst charakterisiert.

Die Ergebnisse dieser Arbeit unterstreichen die Notwendigkeit, die Art und Weise zu überdenken, in der Stoffwechselwege bei Vorhandensein von Enzympromiskuität vorhergesagt werden.

English language declaration

27 February 2020

English language declaration

I, Fabian Kovacs, as a native English speaker, hereby declare the dissertation "Enzymatic promiscuity at the origin of metallo-β-lactamases and within the α/β-hydrolase superfamily" written by Pablo Pérez-García to be linguistically sound.

Fabian Kovacs

Hamburg,

Declaration on oath / Eidesstattliche Versicherung

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Hamburg, 06.03.2020

Pablo Pérez-García

List of abbreviations

Å	Ångström
A. ac.	Amino Acid
AB	Antibiotic
ABH	α/β-Hydrolase
Abs.	Absorbance
AF	Antifoam
AHL	Acyl Homoserine Lactone
Amp	Ampicillin
AOX1	Alcohol Oxidase 1
Approx.	Approximately
ASU	Asimmetric Unit (cryst.)
B	B Form of Atomic Displacement Parameter (cryst.)
B.	*Bacillus*
Bf	Buffer
bis-*p*NPP	Bis(*p*NP)-phosphate
BLASTP	Protein-Protein Basic Local Alignmen Search Tool
bp	Basepair
BRENDA	BRaunschweig ENzyme DAtabase
BYGM	Buffered extra-YNB Glycerol Methanol
C2	Acetate
C4	Butyrate
C8	Caprylate
C12	Laurate
C14	Mystirate
C16	Palmitate
C18	Stearate
CC	Pearson's correlation coefficient (cryst.)
CD	Conserved Domain
CEC	Cefaclor
CFT	Ceftibuten
Cm	Chloramphenicol
c-Myc	Cellular-Myelocytomatosis (Epitope-Tag)
CPSF	Cleavage and Polyadenylation Specificity Factor
Cryst.	Crystallography Term
CTX	Cefotaxime
D.	*Desulfurococcus*
Da	Dalton
DMF	Dimethylformamide
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
dsDNA	Double Stranded DNA
DUF	Domain of Unknown Function
E.	*Escherichia*
eDNA	Environmental DNA
e.g.	*exempli gratia* ("for example")
EC	Enzyme Comission
EPDM	Ethylene Propylene Diene Monomer
EPPS	3-[4-(2-Hydroxyethyl)piperazin-1-yl]propane-1-sulfonic acid
ESTHER	ESTerases and alpha/β-Hydrolase Enzymes and Relatives
etc.	*et cetera* ("and the rest")
EtOH	Ethanol
e-value	Expected Value
g	Gramm
GO	Gene Ontology
GOLD	Genomes OnLine Database
H.	*Homo*
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His_6	Hexahistidin-Tag
I	Intensity (cryst.)
I.	*Ignicoccus*
i.e.	*id est* ("that is")
ICP-MS	Inductively Coupled Plasma-Mass Spectrometry
IMG	Integrated Microbial Genomes and Microbiomes
IPM	Ipenem
IPTG	Isopropyl-β-d-1-thiogalactopyranoside
ITC	Isothiocyanate
k	Kilo (10^3)
K	Kelvin
K.	*Komagatella*
k_{cat}	Catalytic Rate Constant
KH	K Homology
K_M	Michaelis Constant
L	Liter
LB	Lysogenic Broth
LOR	Loracarbef
m	Milli (10^{-3})

M	Molar / Mega (10^6)
M.	*Methanolobus / Methanosarcinia*
MA	Cefamandole
MβL	Metallo-β-Lactamase
MetOH	Methanol
MEZ	Mezlocillin
MGY	Minimal Glycerol
min	Minute
ML	Maximum-Likelihood
MM	Minimal Methanol
mol	Mole
MSA	Multiple Sequence Alignment
MSM	Mineral Salt Medium
MW	Molecular Weight
n	Nano (10^{-9})
N.	*Nanoarchaeum*
nanoDSF	Nano Differential Scanning Fluorimetry
NAPE	N-Acylphosphatidylethanolamine
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
nr	Non-Redundant
OD_{600}	Optical Density at 600 nm
ORF	Open Reading Frame
p	Pico (10^{-12})
P.	*Pichia / Pyrodictium / Pyrococcus / Pectobacterium*
PAGE	PolyAcrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Database
PDE	Phosphodiesterase
PEG	Polyethylene Glycol
Pfam	Protein Family
PLC	Phospholipase C
PLD	Phospholipase D
PMSF	Phenylmethylsulfonyl Fluoride
***p*NP**	*para*-Nitrophenyl
***p*NPP**	*p*NP-Phosphate
***p*NPPC**	*p*NP-Phosphatidylcholine
***p*NPPP**	*p*NP-Phenylphosphonate
pO_2	Partial Pressure of Oxygen
PTE	Phosphotriesterase
PVR	Protein Variable Region
R	Agreement Index (cryst.)
R^2	Coefficient of determination
RMS	Root-Mean-Square (cryst.)
RMSD	Root-Mean-Square Deviation
RNA	Ribonucleic Acid
s	Second
S	Substrate
S.	*Staphylothermus / Streptomyces*
SAD	Single-wavelength Anomalous Diffraction (cryst.)
SB3-14	Myristyl Sulfobetaine
SDS	Sodium Dodecyl Sulfate
ssDNA	Single Stranded DNA
T	Temperature
T.	*Thermophilum / Thermotoga*
T_{ann}	Annealing Temperature
TBT	Tributyrin
TEM	Transmission Electron Microscopy
T_m	Melting Temperature
Tris	Tris(hydroxymethyl)aminomethane
U	Unit
UlaG	Utilization of L-Ascorbate Protein G
V	Volume
V.	*Vibrio*
v_{max}	Maximum Rate
WT	Wildtype
WW	Wet Weight
w/w	Weight by Weight
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
YNB	Yeast Nitrogen Base
Zeo	Zeocin™
ZOI	Zone Of Inhibition
β-CASP	MβL, CPSF, Artemis, Snm1, Pso2
ε	Extintion Coefficient (M^{-1} cm^{-1})
µ	Micro (10^{-6})

List of figures

List of tables

List of publications

Some of the results obtained from the work on this thesis resulted in the following publications (or manuscripts under revision at the time of submission):

- Martinez-Martinez, M., C. Coscolin, G. Santiago, J. Chow, P. J. Stogios, R. Bargiela, C. Gertler, J. Navarro-Fernandez, A. Bollinger, S. Thies, C. Mendez-Garcia, A. Popovic, G. Brown, T. N. Chernikova, A. Garcia-Moyano, G. E. K. Bjerga, **P. Perez-Garcia**, T. Hai, M. V. Del Pozo, R. Stokke, I. H. Steen, H. Cui, X. Xu, B. P. Nocek, M. Alcaide, M. Distaso, V. Mesa, A. I. Pelaez, J. Sanchez, P. C. F. Buchholz, J. Pleiss, A. Fernandez-Guerra, F. O. Glockner, O. V. Golyshina, M. M. Yakimov, A. Savchenko, K. E. Jaeger, A. F. Yakunin, W. R. Streit, P. N. Golyshin, V. Guallar, M. Ferrer and The INMARE Consortium (2018). "Determinants and Prediction of Esterase Substrate Promiscuity Patterns." ACS Chem Biol 13(1): 225-234.
- Kobus, S., **P. Perez-Garcia**, A. Hoeppner, N. Holzscheck, F. Kovacic, W. R. Streit, K. E. Jaeger, J. Chow and S. H. J. Smits (2019). "Igni18, a novel metallo-hydrolase from the hyperthermophilic archaeon *Ignicoccus hospitalis* KIN4/I: cloning, expression, purification and X-ray analysis." Acta Crystallogr F Struct Biol Commun 75(Pt 4): 307-311.
- **P. Perez-Garcia**, S. Kobus, A. Hoeppner, N. Holzscheck, C. H. Strunk, H. Huber, K.-E. Jaeger, F. Kovacic, S. Smits, W. R. Streit, Jennifer Chow (2020) “The origin of metallo-beta-lactamases: The highly promiscuous hydrolase Igni18 from the Crenarchaeon *Ignicoccus hospitalis* represents the paradigm of an ancestral enzyme.” (submitted)

Furthermore, these and other results obtained during my work at the research group of Prof. Dr. Wolfgang R. Streit have also been presented by me in the following conferences:

- **Pérez-García, P.**, Kovacic, F., Smits, S.H.J., Holzscheck, N., Bernhardt, C., Jäger, K.-E., Chow, J., Streit, W.R. (2018) "Heterologous expression of a hyperthermophilic lipase from *Ignicoccus hospitalis* KIN4/I [Crenarchaeota] in *Pichia pastoris*" Annual Conference of the Association for General and Applied Microbiology (VAAM) – Wolfsburg 2018
- **Pérez-García, P.**, Janus, M., Kinfu, B.M., Kovacic, F., Smits, S.H.J., Holzscheck, N., Schwaneberg, U., Jäger, K.E., Chow, J., Streit W.R. (2018) "A combined *in vitro* expression system for cell-free screening and characterization of proteins from non-cultivated organisms and hyperthermophilic archaea" 9th International Congress on Biocatalysis (Biocat) – Hamburg 2018
- **Pérez-García, P.**, Hüpeden, J., Güllert, S., Spieck, E., Chow, J., Streit, W.R. (2019) "Unveiling the biotechnological potential of Recirculating Aquaculture Systems (RAS) through metagenomics and metatranscriptomics" Annual Conference of the Association for General and Applied Microbiology (VAAM) – Mainz 2019

Table of contents

I. INTRODUCTION

I. Introduction

1.1 Genomic annotation *vs.* "dark matter"

Proteins only from a closely related group, of which at least a representative has been characterized experimentally, can be accepted to be correctly annotated. However, during the last 15 years sequencing costs have dropped drastically, mainly due to the development of high-throughput next-generation sequencing (NGS) systems which will soon be able of producing up to 300 Gbp in a single run (*e.g.* NextSeq 2000 by Illumina®, San Diego, CA, USA). As a result, thousands of complete genomes are now publicly available in online databases like GOLD (Mukherjee *et al.* 2019), which houses 342,492 sequencing projects and the genomes of 375,419 organisms at the time of this thesis' publication (Figure 1).

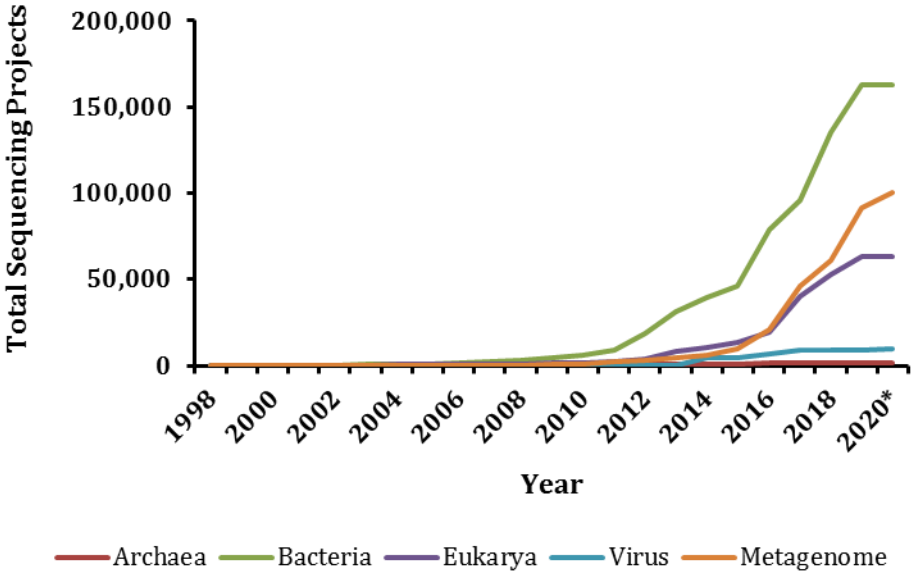


Figure 1: Sequencing projects in the GOLD database – yearly evolution.
Data was extracted from https://gold.jgi.doe.gov/statistics (February 2020). *Values for the year 2020 only contemplate the number of projects uploaded during the first two months of the year.

This number is elevated into the billions when converted to putative protein-coding genes. A lot of work on functional characterization of proteins has already been done, especially on a subset of model organisms which are usually 1) genetically accessible, 2) representative of a phylogenetic group, 3) easy to cultivate or 4) result from special interest because of economical, medical or other reasons [*e.g. Escherichia coli*, *Bacillus subtilis*, *Methanosarcina acetivorans*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Zea mays*, *Caenorhabditis elegans*, *Drosophila melanogaster* or *Mus musculus*; (Leigh *et al.* 2011, Leonelli and Ankeny 2013, Blount 2015)]. Although many advances have been made in the characterization and manual annotation of proteins related to primary and secondary metabolism and their regulation, a substantial percentage of coding genes and non-coding DNA remains, to the present day, without an assigned molecular and/or biological function. This fraction of a genome is referred to as "dark matter" (Ellens *et al.* 2017, Makarova *et al.* 2019). It is estimated that 20 % of the proteins encoded by *E. coli* K12 and 25 % by *A. thaliana* Col-0 ecotype still do not have any function prediction (IMG and EnsemblPlants, February 2020).

To reduce the uncharacterized genomic regions, annotation gained from functional characterization is transferred to genes from which a certain homology can be inferred. These methods often lead to inaccurate or possibly incorrect predictions. In fact, misannotation in public databases has been observed for many protein families, error propagation being suggested as a primary cause (Nobre *et al.* 2016, Lockwood *et al.* 2019). A proper annotation should contain information about a catalytic reaction as well as a role in a metabolic pathway (Ellens *et al.* 2017). Unfortunately, some genes appear incompletely annotated, without a defined role in the cell [*e.g.* '(metallo/metal-dependent-) hydrolase' or 'phosphatase' without a link to a specific reaction]. This poor annotation is not strictly "dark

matter" but can be referred to as "grey matter". Additionally, traditional annotation hardly contemplates the possibility of proteins displaying enzyme promiscuity.

1.2 Enzyme promiscuity as an ancestral trait

Enzymatic or substrate promiscuity is the ability of an enzyme to catalyze a broad substrate spectrum, usually accepting large substrates (Martinez-Martinez *et al.* 2018). This feature is important from evolutionary, environmental, and structural points of view - but has some biotechnological implications as well. Enzyme promiscuity is believed to be an ancient trait which would get lost as the protein evolves from lower to higher substrate specificity (Braakman and Smith 2014, Wheeler *et al.* 2016). Organisms with small, compact genomes - usually deep-branching - tend to harbor a large proportion of genes coding for promiscuous enzymes. These are energetically more favorable than specialized enzymes since the cell does not require many different proteins to take up different substrates (Giovannoni *et al.* 2014, Price and Wilson 2014). Enzymes from an early point of evolution can serve as starting points for protein engineering approaches (Sanchez-Ruiz 2017). Finally, promiscuous enzymes are highly appreciated for biotechnological conversions since they can be applied to many processes, thus reducing production costs and process development times for multiple enzymes (Schmid *et al.* 2001, Ferrer *et al.* 2015).

It is believed that promiscuity might be related to protein flexibility, the existence or absence of structural elements (*e.g.* lid domains) or alterations in the proximity of the active-site cavity and/or access channels (Holmquist 2000). A possible general explanation

for an enzyme displaying substrate promiscuity involving the 'active site effective volume' was proposed recently (Martinez-Martinez *et al.* 2018).

The metallo-β-lactamase (MβL) and α/β-hydrolase (ABH) protein families include mostly enzymes with an extraordinary diversity. Its ancient, less evolved members are good candidates to display a remarkable promiscuity.

1.3 The enzymatic diversity of metallo-β-lactamases and α/β-hydrolases

The MβL-fold consists of a four-layered β-sandwich with two mixed β-sheets flanked by α-helixes (Figure 2), with the metal-binding site(s) located at one edge of the β-sandwich (Callebaut *et al.* 2002). The MβL superfamily includes mainly hydrolytic enzymes carrying out a large array of biological functions. The Class B metallo-β-lactamases (MβLs), giving name to this protein fold, perform only a minor portion of the activities comprised in this superfamily. Nevertheless, continuous outbreaks of MβL-producing multi-resistant pathogens make this activity a major threat to humankind (Falcone *et al.* 2020). Some of the enzymatic activities known for MβLs are summarized in Table 1.

MβLs can also be found fused to transporter proteins like the natural competence protein ComEC, comprised of a domain of unknown function (DUF4131) believed to bind DNA, a transmembrane competence domain for uptake of ssDNA and a MβL domain that would degrade a single strand of the imported dsDNA (Pimentel and Zhang 2018).

The variety of biological roles is, amongst others, reflected in distinctions in the metallo-chemistry of the active pocket. For instance, classical members of the superfamily are di-, or

less commonly mono-, Zn-dependent hydrolases. Nevertheless, MβLs exist with other metal requirements for their activity (Garces *et al.* 2010, Makris *et al.* 2013).

Table 1: Metallo-β-lactamases and their activities.
Only the main activities of each enzyme are shown. *Cleavage and Polyadenylation Specificity Factor **Phosphodiesterase. ***N-acetylphosphatidylethanolamine-hydrolysing phospholipase D. ****Utilization of L-Ascorbate protein G.

Enyzme name	EC no.	Function
AHLase	3.1.1.81	Hydrolysis and opening of the homoserine lactone rings of acyl homoserine lactones (AHLs).
Aryldialkylphosphatase	3.1.8.1	Phosphotriesterase (PTE): degradation of paraoxon, parathion and other phosphotriesters.
Arylsulfatase	3.1.6.1	Hydrolysis of a wide range of sulfate groups (*e.g.* from phenol sulfates).
β-hydroxylase	1.14.99.65	β-hydroxylation of 4-amino-L-phenylalanine (L-PAPA) to form L-p-aminophenylserine.
Class B β-lactamase	3.5.2.6	Hydrolysis of a wide range of β-lactam antibiotics.
CPSF*	3.1.13.B1	Exonucleolytic cleavage in the 5'- to 3'-direction to yield nucleoside 5'-phosphates.
-"-	3.1.4.1	Removal of 5'-nucleotides successively from the 3'-hydroxy termini of 3'-hydroxy-terminated oligonucleotides.
Cyclic nucleotide PDE**	3.1.4.17	Hydrolysis of nucleoside 3',5'-cyclic phosphate (cAMP/cGMP) to nucleoside 5'-phosphate.
Flavodiiron proteins	1.18.1.4	Reduction of oxygen and/or nitric oxide to water or nitrous oxide.
Glyoxalase II	3.1.2.6	Hydrolysis of S-D- lactoylglutathione to D-lactate.
ITCase	3.5.5.8	Hydrolysis of an isothiocyanate to an amine and carbonyl sulfide.
NAPE-PLD***	3.1.4.54	Hydrolysis of NAPE to N-acylethanolamine and a 1,2-diacylglycerol 3-phosphate.
Persulfide dioxygenase	1.13.11.18	Oxidation of glutathione persulfide to glutathione and persulfite (in the mitochondria).
Phosphodiesterase I	3.1.4.1	5'-exonuclease.
Phospholipase D	3.1.4.4	Hydrolysis of a phosphatidylcholine to choline and a phosphatidate.
RNase J	4.6.1.22	5' end-independent endonuclease.
-"-	3.1.4.1	Removal of 5'-nucleotides successively from the 3'-hydroxy termini of 3'-hydroxy-terminated oligonucleotides.
tRNase Z	3.1.26.11	Removal of extra 3' nucleotides from tRNA precursors.
UlaG****	3.1.1.-	Hydrolysis of L-ascorbate-6-phosphate to 3-keto-L-gulonate-6-phosphate.

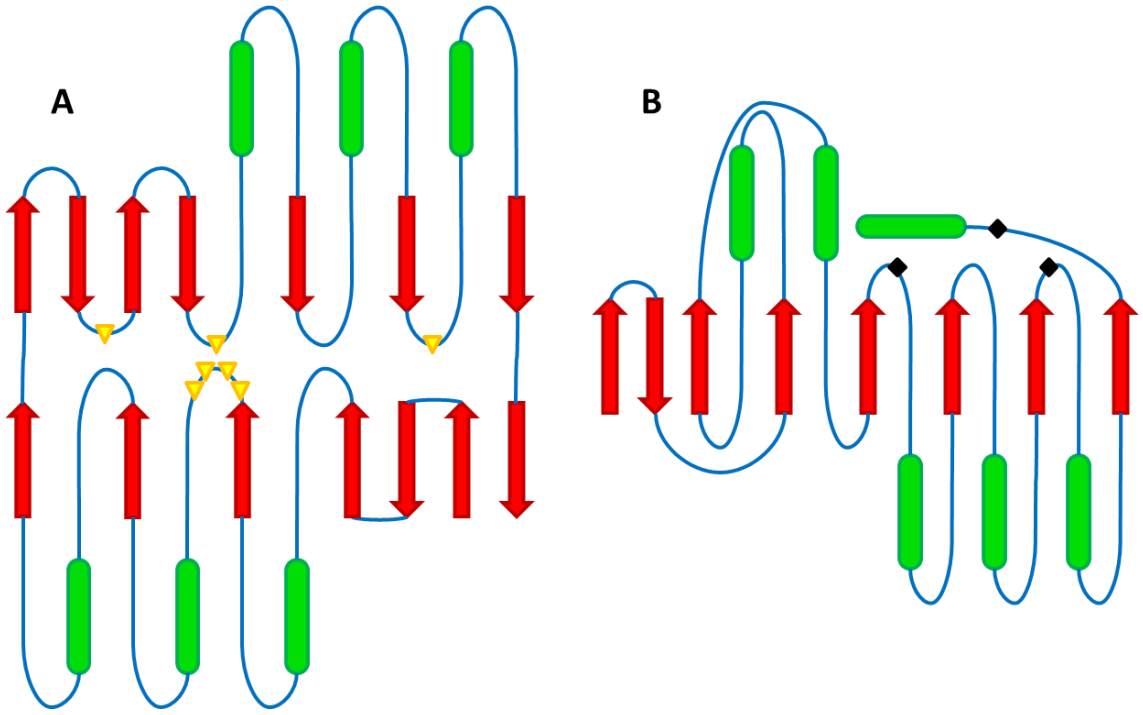


Figure 2: The MβL-fold and the ABH-fold.
Depicted are the Igni18-like core domain of the MβL family (A) and the prototypic ABH-fold (B) as in (Ollis *et al.* 1992). Red arrows represent β-strands and green blocks α-helixes. Yellow inverted triangles mark the seven amino-acid residues necessary for coordination of two Zn ions and catalytic activity. Black squares highlight the position of catalytic-triad residues.

The α/β-hydrolase (ABH) fold consists of a very conserved structure (Figure 2), formed by an α/β sheet, not barrel, of eight β-sheets connected by α-helices (Ollis *et al.* 1992). The enzymes all have a Nucleophile-His-Acid catalytic triad with its residues positioned in loops that are the best conserved of the fold (Holmquist 2000). Members of this family perform very diverse enzymatic activities, including acetylcholinesterase (EC 3.1.1.7), dienelactone hydrolase (EC 3.1.1.45), lipase (EC 3.1.1.23), thioesterase (EC 3.1.2), serine carboxypeptidase (EC 3.4.16.1), proline iminopeptidase (EC 3.4.1.4), haloalkane dehalogenase (EC 3.8.1.5), haloperoxidase (EC 1.11.2.1), epoxide hydrolase (EC 3.3.2.3), hydroxynitrile lyase (EC 4.1.2.37) and many others (Holmquist 2000, Marchot and Chatonnet 2012). Some of them bind metals [*e.g.* carboxypeptidase A, (Ollis *et al.* 1992)].

Interestingly, some members of this family do not catalyze any reaction, but rather have other cellular functions such as hormone precursors or transporters, chaperones or routers of other proteins (Carr and Ollis 2009, Lenfant *et al.* 2013). According to the ESTHER database (Lenfant *et al.* 2013), there are at least 215 subfamilies of ABHs. It is believed that all ABHs have diverged from a common ancestor (Ollis *et al.* 1992).

Both enzyme families are ubiquitously distributed within Eukarya, Bacteria and Archaea - which indicates an ancient origin of these protein-folds (Daiyasu *et al.* 2001, Bebrone 2007, Marchot and Chatonnet 2012, Keshri *et al.* 2018). For this reason, and for their adaptations and capability to perform all the mentioned activities, the MβL- and the ABH-folds can be considered as evolutionary successful. In addition, enzymes from these families that did not yet develop a specialized activity could display substrate promiscuity, a feature an organism like *Ignicoccus hospitalis* could take for its advantage.

1.4 *Ignicoccus hospitalis*: small genome, "dark matter", MβLs and ABHs

Ignicoccus hospitalis is a hyperthermophilic Crenarchaeon, isolated from a low marine hydrothermal system at the Kolbeinsey Ridge, to the north of Iceland (Paper *et al.* 2007). It grows chemolithoautotrophically at 90 °C under strictly anaerobic conditions by reducing sulfur with H_2 and fixating CO_2 via a newly discovered dicarboxylate/4-hydroxybutyrate assimilation pathway (Huber *et al.* 2008). The species name "*hospitalis*" is due to its association with *Nanoarchaeum equitans* (Figure 3), forming the first Archaeon-Archaeon symbiotic relation ever observed to occur naturally (Wrede *et al.* 2012). Due to the extremely reduced genome of *N. equitans* (only 490 Kbp), it completely relies on the supply

of essential biological macromolecules from its host [*e.g.* amino acids and membrane lipids; (Waters *et al.* 2003)].

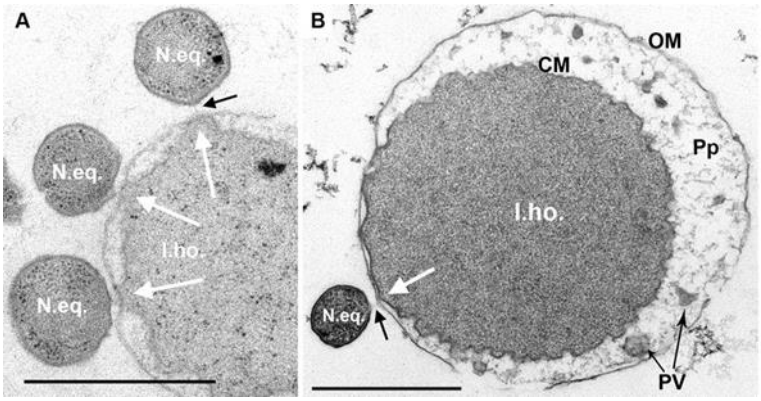


Figure 3: *I. hospitalis* in association with *N. equitans*.
TEM. I.ho., *I. hospitalis* cell; CM, cytoplasmic membrane; OM, outer membrane; Pp, periplasm; PV, periplasmic vesicles; N.eq., *N. equitans* cell. White arrows point to the contact site where the *I. hospitalis* outer membrane is in close contact with the cytoplasmic membrane. Black arrows, fibrous material in the gap between the two cells. Bars, 1 µm. Figure extracted from (Jahn *et al.* 2008).

Surprisingly, the genome of *I. hospitalis*, is not large either. Coding for only 1,496 genes and with a size of only 1.3 Mbp (IMG), it is one of the smallest known for a free-living organism (Podar *et al.* 2008). Small genomes are usually tightly and efficiently packed and do not allow for much regulation. In fact, approx. 97 % of the predicted genes had a transcript and approx. 80 % of the coding proteins were present in *I. hospitalis* cells grown under laboratory conditions (Giannone *et al.* 2015).

At least 12 predicted MβLs and 94 different putative hydrolases are encoded in *I. hospitalis*' genome, but none of them has been studied in detail. Many among its hydrolases are probably binding metals, as they are annotated as metallo-hydrolases. Metal-dependent enzymes require the binding of ions like Mg^{2+}, Ca^{2+}, Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+} or Zn^{2+} for their functionality. *I. hospitalis* possesses transporters for metals like Ni^{2+} and Fe^{2+} (NiFe-

hydrogenases), Mg^{2+}, Co^{2+} and putative ABC transporters for Zn^{2+} and Mn^{2+} located across the outer and inner membrane and the intermembrane compartment (Giannone *et al.* 2011, Giannone *et al.* 2015).

Almost 40 % of the genes encoded by the genome of *I. hospitalis* lack a function prediction and remain to date as "dark matter" (IMG, February 2020), which could be the key to elucidate how this organism can live with such a limited enzymatic repertoire. A method to shed light on unknown genes or specifically search for activities is metagenomics.

1.5 Metagenomics: 'enlightening' the "dark and grey matter"

It is widely accepted that 99.8 % of the microorganisms present in many environments are not cultivable under laboratory conditions, which makes genomic "dark and grey matter" way more immense. Metagenomic approaches originate as an opportunity to access environmental DNA and find new enzymes in the "microbial dark matter" without the need to cultivate the specific microbes (Streit and Schmitz 2004). To date, two different approaches can be followed separately, but they can complement each other (Figure 4). Function-based metagenomic analyses start with the isolation of environmental DNA (eDNA; *e.g.* from sediments, dung or intestinal tracts of animals, water/air filtrates, *etc.*) and the creation of a metagenomic library using a vector of choice (*e.g.* cosmids, fosmids) and a host strain, usually *E. coli* (Streit and Schmitz 2004, Schmeisser *et al.* 2007). Once the library is constructed, it can be screened for specific activities against different substrates using a suitable method. Sequence-based metagenomic approaches part either from the same sampled material which is then sequenced (NGS) or from data available in public

domain resources such as NCBI (www.ncbi.nlm.nih.gov) or IMG (www.img.jgi.doe.gov). Metagenomes can be screened for enzymes of interest via bioinformatics methods such as Hidden Markov Model (HMM) searches (Mistry *et al.* 2013).

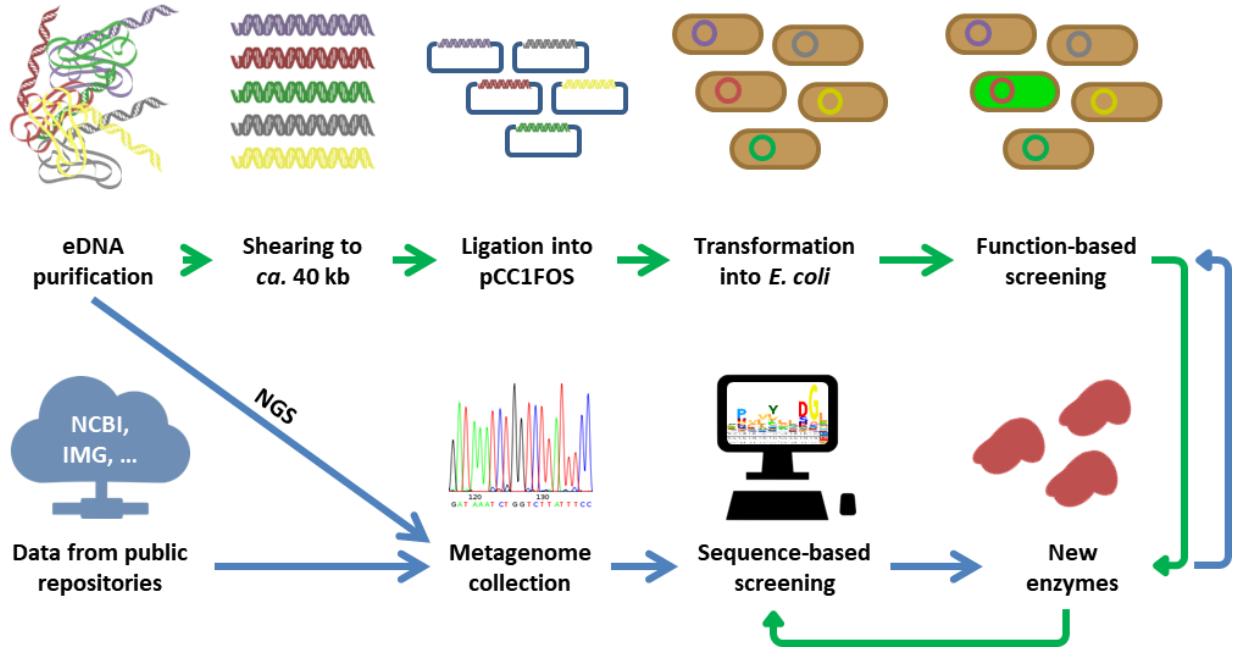


Figure 4: Function- and sequence-based metagenomic screening methods.
Green arrows represent a function-based approach and blue arrows the sequence-based approach using bioinformatics. Both methods can complement each other. pCC1FOS: commonly used fosmid for library preparation; NGS: Next-Generation Sequencing.

Most metagenomic screenings aim for enzymes which have value for industrial use. Lipases (ABHs), together with aldo-keto-reductases, β-transaminases and alcohol dehydrogenases are among the main targets (Martínez-Martínez *et al.* 2016).

During previous works in the group of Prof. Dr. Wolfgang R. Streit, several metagenomic libraries were screened for activity on tributyrin, a short-chained (C4) tri-ester of glycerol. All positive fosmid-/cosmid-carrying clones able to degrade the model substrate, along with other clones also displaying esterase/lipase activity on different *p*NP-esters were collected

in the so called "lipase toolbox" - an assortment of putative, new, still uncharacterized enzymes with a probable biotechnological application. The origins of all active clones included in the "lipase toolbox" are summarized in Table 2.

Table 2: Origin of the metagenome-derived clones displaying esterase activity.
All metagenomic libraries were constructed at the lab of Prof. Dr. Streit. *Ethylene Propylene Diene Monomer.

Origin	No. of clones screened	No. of active clones identified
Thermus sp. enrichments (Chow 2008)	600	6
Thermophilic soil enrichment I (Köhler 2007)	2,600	3
Thermophilic soil enrichment II (Chow 2008)	6,300	5
EPDM*-biofilm (Schmeisser *et al.* 2003)	1,440	7
Elephant feces (Ilmberger *et al.* 2014)	20,000	72
Teufelsbrück, river Elbe sediment (Krohn 2010)	10,080	44
Glückstadt, river Elbe sediment (Böhnke 2010)	40,000	5
TOTAL	**81,020**	**142**

1.6 Aim of the study

This work focuses on enzyme promiscuity, its origin and characteristics of promiscuous enzymes within two protein families. The thesis is divided in two parts:

The first part of my thesis examined the promiscuous origin of the MβL family with respect to the very ancient archaeal enzymes. Therefore, I investigated the catalytic activities and the structural properties of the predicted ancestral MβL family enzyme Igni18. The protein, from the crenarchaeon *I. hospitalis*, was annotated as putative MβL. This annotation can only be considered as a general function prediction since it cannot be linked to a certain function within the organism. Thus, the structural, biochemical and catalytic investigations in this work contribute to the in-depth functional characterization of this novel and hitherto

not characterized protein in *I. hospitalis*. Since Igni18 belongs to the "dark/grey matter" proteins, this thesis contributes to the elucidation and general functional assignment within the archaeal phylum. The project started with the optimization of heterologous expression of *igni18* to produce enough recombinant protein for crystallization experiments. Further structural analysis of Igni18 and other MβLs in combination with an exhaustive biochemical characterization aimed at identifying traits leading to promiscuity in this protein family, as well as discovering evolutionary keys within this very important enzyme family.

During the second part of this study, I focused on enzyme promiscuity within the ABHs on a larger scale. Starting from a collection of 142 active, metagenome-derived putative esterases - the "lipase toolbox" - their ability to degrade esters from a large substrate collection should be assayed. A combined structural, biochemical and catalytic profiling should help to identify specific, moderate- and highly promiscuous enzymes. These results, combined with data from collaboration partners, should give enough evidence to discover a general pattern from which specificity or promiscuity of an enzyme could be predicted *in silico*. Furthermore, single enzymes should be sub-cloned and characterized for their possible implementation in biotechnological processes.

II. MATERIALS AND METHODS

II. Materials and methods

2.1 Microbial strains and cultivation

During these studies, both bacterial and eukaryotic cloning and/or expression systems were used. A list of the base strains referred to in this study is presented in Table 3.

Table 3: Bacterial and yeast strains used in this study.

Strain	Phenotype	Source
E. coli EPI100	F- *mcr*A Δ(*mrr-hsdRMS-mcrBC*) Φ80*dlacZ*ΔM15 Δ*lacX*74 *recA1 endA1 ar-aD139* Δ(*ara, leu*) 7697 *galU galK* λ-*rpsL nupG*	Epicentre (Chicago, IL, USA)
E. coli EPI300™-T1R	F– mcrA Δ(mrr-hsdRMS-mcrBC) (StrR) φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ– rpsL nupG trfA tonA dhfr	Epicentre (Chicago, IL, USA)
E. coli VCS257	Derivative of DP50 *supFc*	Agilent Technologies (Santa Clara, CA, USA)
E. coli XL-1 blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F′ proAB lacIq ZΔM15 Tn10 (Tetr)]	Agilent Technologies
E. coli DH5α	F- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^- m_K^+$), λ^-	Invitrogen (Karlsruhe, Germany)
E. coli BL21 (DE3)	F– *ompT gal dcm lon hsdS*$_B$($r_B^- m_B^-$) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^+$]$_{K-12}$(λ^S)	Merck (Darmstadt, Ger-many)
E. coli Shuffle® T7	F′ *lac*, *pro*, *lacIq / Δ(ara-leu)7697 araD139 fhuA2 lacZ::T7 gene1 Δ(phoA)PvuII phoR ahpC* galE (or U) galK λatt*::pNEB3-r1-*cDsbC* (SpecR, *lacIq*) *ΔtrxB rpsL150*(StrR) *Δgor Δ(malF)3*	New England Biolabs (Frankfurt am Main, Germany)
P. pastoris X-33	WT	Thermo Fisher Scientific (Waltham, MA, USA)

All media and buffers used in this study were autoclaved for 20 min at 121 °C. Bacterial strains (*E. coli*) were usually cultivated at 37 °C under appropriate aerobic conditions in LB medium [in g/L: yeast extract 5, peptone 10, NaCl 10 (Bertani 1951)] containing antibiotics (Table 4) unless otherwise indicated. The yeast *P. pastoris* was grown at 28-30 °C in the media YPD (in g/L: yeast extract 10, peptone 20, dextrose 20) or Minimal Glycerol (MGY; 10 % glycerol, 1x YNB) as indicated in the EasySelect™ Pichia Expression Kit (Thermo Fisher Scientific, Waltham, MA, USA). All media could be prepared as solid plates by adding 12 or 15 g/L of agar-agar (for bacteria or yeast, respectively) prior to sterilization.

Table 4: Antibiotics and supplements.
*For pCC1FOS or pTZ19R, respectively. **For bacteria, yeast and yeast multi-insertion selection, respectively.

Antibiotic/supplement	Solvent	Stock	Working concentration
Ampicillin (Amp)	70 % EtOH	100 mg/mL	100 µg/mL
Chloramphenicol (Cm)	EtOH	25 mg/mL	12.5 or 25 µg/mL*
Zeocin™ (Zeo)	H_2O	100 mg/mL	25 or 100 or 1,000 µg/mL**
IPTG	H_2O	100 mg/mL	100 µg/mL
X-Gal	DMF	50 mg/mL	50 µg/mL

2.2 DNA techniques

2.2.1 Next generation sequencing (NGS)

The fosmids/cosmids carrying genes coding for putative esterases were sequenced at the Göttingen Genomics Laboratory (G2L, Göttingen, Germany). Clones were grown overnight in 5 mL LB with the correspondent antibiotic. Fosmid clones (pCC1FOS) were supplemented with 0.01 % arabinose to induce vector replication and its presence in high copy number. After centrifugation, pellets were stored at -20 °C until further use. Fosmid

purification was performed with the Presto™ Mini Plasmid Kit (Geneaid Biotech Ltd., New Taipei City, Taiwan) and samples were diluted in 10 mM Tris pH 8. Concentration, quality and purity of the DNA were assayed with agarose gel electrophoresis and measurements with a Nanodrop 2000 (Thermo Fisher Scientific, Waltham, USA). Samples were pooled into four mixes containing 40 constructs each, with a final volume of 40 µL and a concentration of 2.5 ng/µL. Pools were labeled with "LipH_f_#", where "#" indicated the pool number, and sent to G2L for Illumina MiSeq sequencing. Quality assessment and assembling of the reads were automatically performed by G2L. The contigs were named after the pool name plus the contig number as "LipH_f_#_*c" where "*" is the contig number. Predicted ORFs were named after the contig plus an identification number (N) as such: "LipH_f_#_*c_N". Enzymes were named "#_enzyme_*c_N". Tools applied for ORF prediction and sequence-based screening for α/β-hydrolases, esterases and lipases are enumerated below (see 5.6 Bioinformatics).

2.2.2 Cloning

Note: Cloning of igni18 *into pPICZ-A and transformation into* P. pastoris *X-33 were carried out by Nicholas Holzscheck and are described in detail in (Holzscheck 2016).*

The constructs used and/or generated in this work for propagation or expression of genes were constructed with the backbones of the vectors listed in Table 5. Cloning into the pCC1FOS and the pDrive vectors occurred in a non-directional manner via blunt-end- and U-A-based ligations, respectively, as indicated by the manufacturer manuals. Unidirectional insertion of genes into the pET21a(+) and pPICZ-A plasmids was directed with the addition of two palindromic endonuclease restriction sites (*e.g. Eco*RI/*Not*I, *Nde*I/*Hind*III) to the primers as described in (Green and Sambrook 2012). *E. coli* was heat-shock transformed via standard methods. The EasySelect™ Pichia Expression Kit (Invitrogen, Carlsbad, CA, USA)

was employed for cloning and expression of *igni18*. Insertion of *Mss*I-linearized DNA constructs into yeast cells was performed based on the manufacturer's manual and as in (Pérez-García 2016) and (Holzscheck 2016). Sanger sequencing to assess the correctness of the inserted DNA fragments and other purposes was performed by Eurofins Genomics Germany GmbH (Ebersberg, Germany).

Table 5: Cloning and expression vectors used in this study.

Name	Size [kbp]	Features	Source
SuperCos	7.9	Cosmid vector, Kan^R, Amp^R, *cos*, T7 and T3 promoters	Agilent Technologies (Santa Clara, CA, USA)
pTZ19R	2.9	Cloning vector, Amp^R, T7-*lac*-promoter	Thermo Fisher Scientific (Waltham, MA, USA)
pWE15	8.2	Cosmid vector, Amp^R, Kan^R, $G418^R$, T7 and T3 promoters	Agilent Technolo-gies (Santa Clara, CA, USA)
pCC1FOS	8.1	Fosmid vector, Cm^R, T7 promoter	Epicentre (Madison, WI, USA)
pDrive	3.9	Cloning vector, Amp^R, Kan^R, T7, SP6 and *lac* promoters	Qiagen (Hilden, Germany)
pET21a(+)	5.4	Expression vector, *lacI*, AmpR, T7-*lac*- promoter, C-terminal His_6-tag coding sequence	Novagen/Merck (Darmstadt, Germany)
pPICZ-A	3.3	Expression vector, Zeo^R, AOX1 homology sites, C-terminal *c.myc* apitope and His_6-tag coding sequences	Thermo Fisher Scientific (Waltham, MA, USA)

2.2.3 Polymerase chain reaction

Polymerase chain reaction (PCR) served two purposes in this study: On the one hand, single genes could be amplified out of genomic DNA or other constructs, introducing specific recognition sites to allow cleavage by an endonuclease prior to (sub-) cloning into a vector. On the other hand, insertion of the gene of interest could be verified via colony PCR. Table 6 shows an extract of the primers employed in this work, including all that are commercially available for sequencing or verification purposes. Other primers designed in this study are

listed in Supplementary Table 1. Reactions were pipetted according to Table 7 and run as in Table 8. In case of a PCR repeatedly failing to amplify a unique band or any band at all, a touchdown program was used (Table 9). Touchdown PCR (Don *et al.* 1991) tackles the problem of mispriming by one or both oligonucleotide amplimers. The annealing temperature of the reaction is decreased every cycle for the first half of the reaction, thus increasing the specificity of the priming and giving advantage to the formation of the expected product.

Table 6: Primers used in this study (Pt. 1).
For a table containing the rest of the primers used in this study, refer to Supplementary Table 1.

Name	Sequence	T_m [°C]	Source
pCC1_for	5'-GGATGTGCTGCAAGGCGATT AAGTTGG-3'	71	Epicentre (Madison, WI, USA)
pCC1_rev	5'-CTCGTATGTTGTGTGGAATT GTGAGC-3'	67	Epicentre (Madison, WI, USA)
M13-20_for	5'-GTAAAACGACGGCCAGT-3'	60	Qiagen (Hilden, Germany)
M13_rev	5'-CAGGAAACAGCTATGAC-3'	55	Qiagen (Hilden, Germany)
T7_prom	5'-TAATACGACTCACTATAGGG-3'	53	Eurofins (Elsberg, Germany)
T7_term	5'-CTAGTTATTGCTCAGCGGT-3'	55	Eurofins (Elsberg, Germany)
Igni18_*Eco*RI_for	5'-CCGAGAATTCGACATGGCCAC GGTTAAGCTGACCTAC-3'	74	(Holzscheck 2016)
Igni18_*Not*I_rev	5'-AGCGGCCGCAAAATTCGAAGG TCACCGTCTCC-3'	73	(Holzscheck 2016)
AOX1_for	5'-GACTGGTTCCAATTGACAAGC-3'	62	Thermo Fisher Scientific (Waltham, MA, USA)
AOX1_rev	5'-GCAAATGGCATTCTGACATCC-3'	62	Thermo Fisher Scientific (Waltham, MA, USA)

Table 7: PCR pipetting scheme.
If not indicated otherwise, *Taq*-Polymerase was used for most reactions (DreamTaq, Thermo Scientific). Values in brackets refer to the proof-reading DNA-polymerase Phusion™ High-Fidelity. *For colony PCR, a single colony was picked, and a fraction of the cells was diluted directly in the reaction as template.

Component	**V [**µL]
DNA*	1
10x Taq Buffer (5x HF/GC Buffer)	2.5 (5)
dNTPs, 2 mM each	1
Primers (2 pmol/µL)	1 each
Polymerase	1 (0.25)
H_20	*ad*. 25

Table 8: Standard PCR program.
Values in brackets refer to hot-start, proof-reading DNA-polymerases (*e.g.* Phusion™ High-Fidelity).
*T_{ann}=((Tm1+Tm2)/2)-5

Step	**T [°C]**	**t [min:s]**	
Initial denaturation	95 (98)	5:00 (3:00)	
Denaturation	95 (98)	0:30	
Annealing	T_{ann}*	0:45	32 cycles
Elongation	72	1:00 (0:30)/kbp	
Final elongation	72	5:00	
Final incubation	10	end	

Table 9: Touchdown PCR program.
Values in brackets refer to hot-start, proof-reading DNA-polymerases (*e.g.* Phusion™ High-Fidelity). *For the first 16 cycles, initial T_{ann} is the calculated T_{ann} + 4 °C and decreases 1 °C for every cycle to a final T_{ann} – 4 °C.

Step	**T [°C]**	**t [min:s]**	
Initial denaturation	95 (98)	5:00 (3:00)	
Denaturation	95 (98)	0:30	
Annealing	T_{ann}*	0:45	16 cycles
Elongation	72	1:00 (0:30)/kbp	
Denaturation	95 (98)	0:30	
Annealing	T_{ann}	00:45	16 cycles
Elongation	72	01:00 (00:30)/kbp	
Final elongation	72	05:00	
Final incubation	10	end	

2.3 Protein production methods

2.3.1 Heterologous protein production in yeast

Fermentation at 10 L scale and HPLC analysis were carried out under supervision of Clemens Bernhardt (Department of Food Microbiology, Universität Hamburg).

Production of Igni18 was performed at 30 °C in buffered extra-YNB glycerol MetOH (BYGM) auto induction medium (Lee *et al.* 2017) without antibiotics for 46 h, as described previously (Holzscheck 2016, Pérez-García 2016, Kobus *et al.* 2019), but the fermentation process was up-scaled to 10 L in an Infors HT Labfors benchtop bioreactor (13 L vessel volume, Infors AG, Bottmingen, Switzerland). The parameters pO_2, T, airflow, pH and addition of NaOH or antifoam were constantly monitored by the system. Samples were taken at several timepoints, filter-sterilized (0.2 µm) and 5 µL were analyzed by HPLC for glycerol and MetOH contents (Bernhardt 2019). Fermentation broth was concentrated by filtration (0.2 µm, Centramate™ 500 S Tangential Flow Filtration System, Pall, Dreieich, Germany) and centrifuged for 10 min at 10,000 rpm (F12-6x500 LEX, Sorvall RC6 Plus, Thermo Scientific, Massachusetts, USA). The pellet was stored at -80 °C until further use.

2.3.2 Protein purification

For protein purification, 15 g pellet corresponding to approx. 1 L of culture were resuspended in 5 mL/g lysis buffer (10 mg/mL SB3-14, 1 mM PMSF, 0.05 M NaH_2PO_4, 0.3 M NaCl, pH 8). Cell disruption and partial purification was accomplished by incubating the cells at 70 °C for 1 h in the presence of the zwitterionic detergent SB3-14 (Zanna and Haeuw 2007). Cell debris was removed from the crude cell extract by centrifugation at 15,000 rpm for 30 min at 4 °C (Sorvall RC6+ centrifuge, SS-34 rotor; Thermo Scientific, Braunschweig, Germany). The clear lysate was loaded onto a Protino® Ni-TED 2000 Packed Column (Macherey-Nagel GmbH & Co, Düren, Germany) and Igni18 was purified according to the

manufacturer's protocol. The eluted protein was concentrated, and metal ions were removed ("stripped") with 25 mM HEPPES-Bf pH 7.5 containing 20 mM EDTA in an ultrafiltration unit (Vivaspin 20, Sartorius AG, Göttingen, Germany). The protein was subsequently dialyzed against 0.1 M potassium phosphate buffer or 5 mM EPPS-Bf pH 7, sterile filtered (0.2 µm) and stored in aliquots at 4 °C for up to several months. Purity and size were determined with SDS-PAGE (Laemmli 1970). Protein concentration was adjusted to 1 or 0.1 mg/mL prior to use.

2.4 Enzymatic assays

2.4.1 Disc-diffusion antibiotic susceptibility test

β-lactamase activity was assayed via a disc-diffusion antibiotic susceptibility test. Mezlocillin 30 µg (MEZ 30), imipenem 10 µg (IPM 10), cefamandole 30 µg (MA 30), loracarbef 30 µg (LOR 30), cefaclor 30 µg (CEC 30), cefotaxime 30 µg (CTX 30) and cefotiam 30 µg (CFT 30) susceptibility discs (Supplementary Figure 1, Thermo Fischer Scientific, Waltham, MA, USA) were incubated overnight at 40 °C with 30 µL 0.1 M potassium phosphate buffer pH 8 containing 1mg/mL stripped protein and 1 mM $NiCl_2$ or $ZnCl_2$. An enzyme-free control was included. The antibiotic susceptibility test was carried out on LB-agar plates and with *E. coli* DH5α cells. After incubation overnight at 37 °C, the zone of inhibition (ZOI) was determined for every antibiotic and condition. The reduction of the ZOI was expressed in percentage.

2.4.2 Activity on model *p*NP-substituted subtrates

The substrates *p*NP-carboxyl esters with various distinct acyl chain lengths (*p*NP-C2 to C18), *p*NP-phosphate (*p*NPP), bis-*p*NP-phosphate (bis-*p*NPP), paraoxon-ethyl and parathion-ethyl were purchased from Sigma-Aldrich (Munich, Germany), *p*NP-phenylphosphonate (*p*NPPP) and *p*NP-phosphorylcholine (*p*NPPC) from Biomol GmbH (Hamburg, Germany, Supplementary Figure 2). 10 mM stock solutions were prepared in 2-propanol and stored at -20 °C. Assays were performed with cell lysate (fosmids clones and sub-cloned esterases) or purified protein (Igni18). Standard assays were performed in 200 µL containing 190 µL Bf with 1 mM substrate and 10 µL enzyme solution (0.1 or 1 mg/mL) or cell lysate from an overnight expression culture (96 deep-well plate) and incubated at 40/60 °C (metagenome-derived enzymes) or 90 °C (Igni18) for 30 min, unless otherwise indicated. Reactions were stopped by the addition of 20 µL 2 M Na_2CO_3 and the formation of *p*-nitrophenolate was measured spectrophotometrically at 405 nm in a Biotek Synergy HT (Bad Friedrichshall, Germany) plate-reader. All assays were performed in triplicate and a buffer control was added to quantify auto-hydrolysis of the substrates. A standard curve with known concentrations of pure *p*-nitrophenolate was used to determine the extinction coefficient (ε) of the hydrolysis product.

For Igni18, the effect of different divalent metal ions (Mg, Ca, Mn, Fe, Co, Ni, Cu and Zn) on the enzyme activity was studied by adding 1 mM of the corresponding metal chloride salts to the reaction. Temperature optimum was determined in the range of 40 to 95 °C. The optimal pH was assayed with different buffers between pH 4 to 10 (0.1 M; pH 4-6: citrate-phosphate buffer; pH 7-8: tris buffer; pH 9-10: carbonate-bicarbonate buffer). For kinetic studies, several substrate concentrations were assayed, and aliquots were taken at different

time points of the reaction and stored on ice until absorbance was measured. One activity U was defined as the amount of protein converting 1 μmol substrate per minute. V_{max}, K_m and k_{cat} were calculated according to Michaelis-Menten kinetics.

2.4.3 "Enzyme fingerprint" on 96 esters

Note: pH-shift assays were performed by collaboration partners at the Institute of Catalysis, CSIC, Madrid, Spain (Mónica Martínez-Martínez, Cristina Coscolín and Manuel Ferrer) with fosmids clones supplied by me.

Prior to enzymatic assays, 500 μL of LB-antibiotic (AB) medium in 2 mL tubes were inoculated with a single colony followed by incubation at 37 °C and 700 rpm in a Thermomixer for 7 h. Then, LB-AB-IPTG agar plates were spread with 300 μL of each culture ensuring a uniform growth. After incubation overnight at 37 °C, 5 mL of 40 mM HEPES-Bf pH 7 were added to each plate. Bacterial cells were detached using sterile disposable Drigalsky spreaders and cellular suspensions were transferred to a 5 mL tube and pelleted by centrifugation at 8000 rpm for 10 min at 4 °C. Pellets were washed twice with 40 mM HEPES-Bf and wet weight (WW) of each culture was recorded. Washed pellets were stored at -80 °C until use. On average, 100 mg/clone (WW) cell pellet was obtained. Enzyme extract solutions were prepared by re-suspending 100 mg (WW) pellet in 0.5 mL of 5 mM EPPS-Bf pH 8 and ester stock solutions were set by dissolving each compound at a concentration of 100 or 10 mg/mL in acetonitrile or DMSO.

Hydrolytic activity was assayed using 96 structurally diverse esters (Supplementary Figure 3) in 384-well plates as previously described (Alcaide *et al.* 2013, Martinez-Martinez *et al.* 2013, Alcaide *et al.* 2015): A 384-well plate was filled with 20 μl of 5 mM EPPS-Bf pH 8 with the use of a QFill3® microplate filler (Genetix Biotech, New Dehli, India). Then, 2 μL of each ester stock solution were added to each well by using a PRIMADIAG liquid handling robot

(EYOWN Technologies SL, Madrid, Spain). Each well was then filled with 20 μL of 5 mM EPPS-Bf pH 8.0 containing 0.45 mM phenol red by using a QFill3® microplate filler. Immediately after, 2 μL of the enzyme extract solution were added to each well. Activity was assayed at 40 °C in a Biotek Synergy HT plate-reader (Bad Friedrichshall, Germany). Hydrolysis of esters was measured spectrophotometrically at 550 nm (ε = 8,450 $M^{-1}cm^{-1}$). All measurements were performed in duplicates and all values were corrected for non-enzymatic transformation. One unit (U) of enzyme activity was defined as the amount of enzyme required to transform 1 μmol of substrate in 1 min under the assay conditions.

2.5 Protein structure determination and characterization

2.5.1 Crystallization

Note: Crystallization experiments and structure resolution were performed by collaboration partners at the Heinrich Heine Universität Düsseldorf (Stefanie Kobus, Astrid Höppner and Sander H. Smits) with protein supplied by me.

Crystallization of Igni18 is described in (Kobus *et al.* 2019). The specific conditions applied to get protein crystals are listed in Supplementary Table 2. The dataset was collected from a single Igni18 crystal on the beamline ID30A-3 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) at -173 °C. Exhaustive methodic details on structure solving, model phasing, ASU definition, manual building and model refinement are available in Pérez-García *et al.* (submitted).

2.5.2 Cofactor analysis

A 12 mg/mL non-stripped Igni18 solution in 50 mM Tris-HCl pH 8 was sent to Analytik Labor Schirmacher GmbH (Hamburg, Germany) for inductively coupled plasma - mass

spectrometry (ICP-MS) analysis. The presence of usual protein-bound divalent metals Mn, Fe, Co, Ni, Cu and Zn was quantified, but also Ti, Se, Nb, Mo, Ag, Cd, Hg and Pb could be potentially detected with this method. A buffer control was included as a blank.

2.6 Bioinformatics

The essential bioinformatics tools and software used for the planning, analysis and general realization of this work are listed in Table 10 and were used according to the manuals for the desired applications.

Table 10: Bioinformatics tools applied in this study.
Only specialized software used for the realization of this work appears in this table. Device-specific software (*e.g.* plate reader) or versions are not listed but can be found at the manufacturer's website.

Name	Application	Source
BLAST (n and p)	Protein and nucleotide sequence searches against the NCBI's nr-database and user-defined databases.	(Boratyn *et al.* 2013)
BioEdit 7.2.1	MSA visualization for Figure 21.	(Hall *et al.* 2011)
BioVinci 1.1.5	Hierarchical clustering analysis and visualization (Figure 34).	BioTuring Inc. (San Diego, Ca, USA)
BRENDA	EC number of enzymatic activities.	(Jeske *et al.* 2019)
CD-search	Detection of conserved domains in protein sequences.	(Marchler-Bauer *et al.* 2017)
COFACTOR	Structure-based molecular function prediction.	(Zhang *et al.* 2017)
ChemMine Tools	Hierarchical clustering of substrate structures.	(Backman *et al.* 2011)
UCSF Chimera 1.14	Protein 3D visualization, structural alignment and PVR analysis.	(Pettersen *et al.* 2004)
ChromasPro 2.1.8	Analysis of Sanger sequencing results.	Technelysium Pty Ltd (South Brisbane, QLD, AU)
CloneManager 9	Design of cloning strategies.	Sci Ed Software (Westminster, CO, USA)
DALI	Heuristic PDB search.	(Holm 2019)
HMMER 3.3	Database search for sequence homologs.	(Mistry *et al.* 2013)
MONA 2	Visualization of compound structures.	(Hilbig and Rarey 2015)

Table 10, cont.

Name	Application	Source
MEGA X (10.0.5)	MSA, molecular evolutionary genetics analysis and visualization of cladograms.	(Kumar *et al.* 2018)
mTM-align	Structural searches against the PDB-database.	(Dong *et al.* 2018)
Pfam 32.0	HMM download and annotation of genes.	(El-Gebali *et al.* 2019)
PhyML 3.3	ML-tree calculation (Figure 35)	(Lefort *et al.* 2017)
Prodigal 2.6.1	ORF discovery.	(Hyatt *et al.* 2010)
Samtools 1.10	Interacting with high-throughput sequencing data (*e.g.* FASTA extraction).	https://github.com/samtools/
T-Coffee	MSA (modes: M-Coffee; PSI-Coffee for Figure 35).	(Notredame *et al.* 2000, Wallace *et al.* 2006)

III. THE PROMISCUOUS ORIGIN OF METALLO-B-LACTAMASES

III. The promiscuous origin of metallo-β-lactamases

In the first part of this thesis, the creanarchaeal MβL Igni18 was structurally, biochemically and catalytically characterized. While enzymes from this family apparently exert a specific function, the results presented here suggest that Igni18 is an ancestral, promiscuous and primordial representative of the MβL family. Since Igni18 is the deepest branching ancestral enzyme functionally characterized, it resembles the paradigm of evolutionary early MβL. Its crystal structure served as a model to describe how an archetype protein could have evolved into modern, specific enzymes. These very interesting and, with respect to enzyme evolution, important results were partly published (Kobus *et al.* 2019) and are in submission at the time of publication of this thesis (Pérez-García *et al.*, submitted).

3.1 Results

3.1.1 Cloning, recombinant protein production and purification

Archaeal genes are often very challenging with respect to their heterologous expression. Attempts to produce native protein in the background of various *E. coli* expression hosts were not successful (data not shown). To overcome this bottleneck, a eukaryotic host was chosen. The yeast *P. pastoris* carrying multiple chromosomally integrated copies of the *igni18* gene was used as an expression system. The gene *igni18* was cloned into the pPICZ-A vector via the restriction sites *Eco*RI and *Not*I, original start codon was modified to ATG and a Kozak sequence (Romanos *et al.* 1992) was added for better expression (Table 6).

At the start of this project, protein production was performed in flasks containing one liter of Minimal Glycerol (MGY) medium, used for initial growth, which was exchanged for Minimal Methanol (MM) medium after 24 h to promote gene expression for another 24 h as in (Pérez-García 2016) and (Holzscheck 2016) and the manual of the Invitrogen™ EasySelect™ *Pichia* Expression Kit (Thermo Fisher Scientific, Waltham, USA). This method yielded approx. 6 mg recombinant, purified Igni18 per liter of culture. In 2017, Lee and colleagues developed the Buffered extra-YNB Glycerol Methanol (BYGM) medium (Lee *et al.* 2017). The medium combines glycerol and MetOH, the growth and the induction carbon source respectively, in a single auto induction medium, thus facilitating the handling and allowing for an easy adaptation of the protocol to a fermenter-based approach. The process was up-scaled to 1.5 L in a benchtop fermenter and a yield of 8 mg/L recombinant and purified Igni18 was obtained. Further upscaling to a total volume of 10 L increased this yield to 12 mg/L (Figure 5).

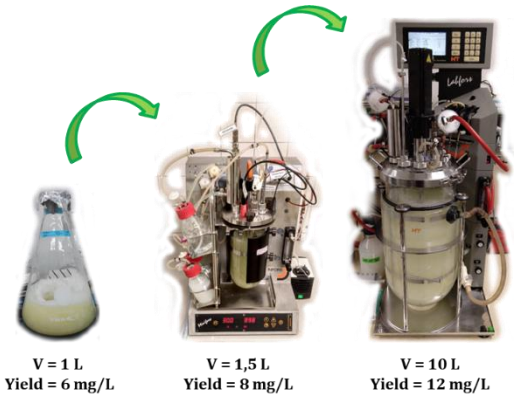


Figure 5: Fermentation upscale to produce Igni18.
Initial protein production in flasks with media swap yielded 6 mg per liter of culture. The use of auto induction medium and a small bioreactor increased the yield to 8 mg/L. Further upscaling of the process improved the production of pure Igni18 to 12 mg/L.

The 13 L fermenter (10 L fermentation broth and 3 L headspace) allowed for automatic recording of the process parameters. A typical fermentation course is represented in Figure 6. During the first ten hours after inoculation, $p0_2$ remained stable at 96-97 % while pH started dropping slightly (Figure 6A), indicative of a lag phase where adaptation to the new medium happened. Then, oxygen uptake by the cells rapidly increased, decreasing $p0_2$ to 78 % and pH dropped to 3.7, indicating metabolic activity of the yeast during the exponential growth phase. Immediately after, pO_2 raised back to 95 % and stayed stable for 6 hours. The pH was automatically set to 5. HPLC analysis of the supernatant after 21 h showed the glycerol concentrations reduced to a minimum and no glycerol could be detected in the medium after 27 h, while MetOH was still present at the initial concentration (Figure 6B). When adaptation of *P. pastoris* to the induction carbon source was completed, the biggest drop in dissolved oxygen was observed (pO_2 40 %, Figure 6A) and the pH dropped again to 3.7. During this time, the AOX1 promotor preceding the *igni18* gene is activated and protein is produced. Consumption of MetOH (41 h, Figure 6B), the only available carbon source and inducer, stops transcription and a hunger peak is observable. Addition of MetOH triggered the immediate metabolic activation of the yeast, demonstrating that the cells were still viable (Figure 6A).

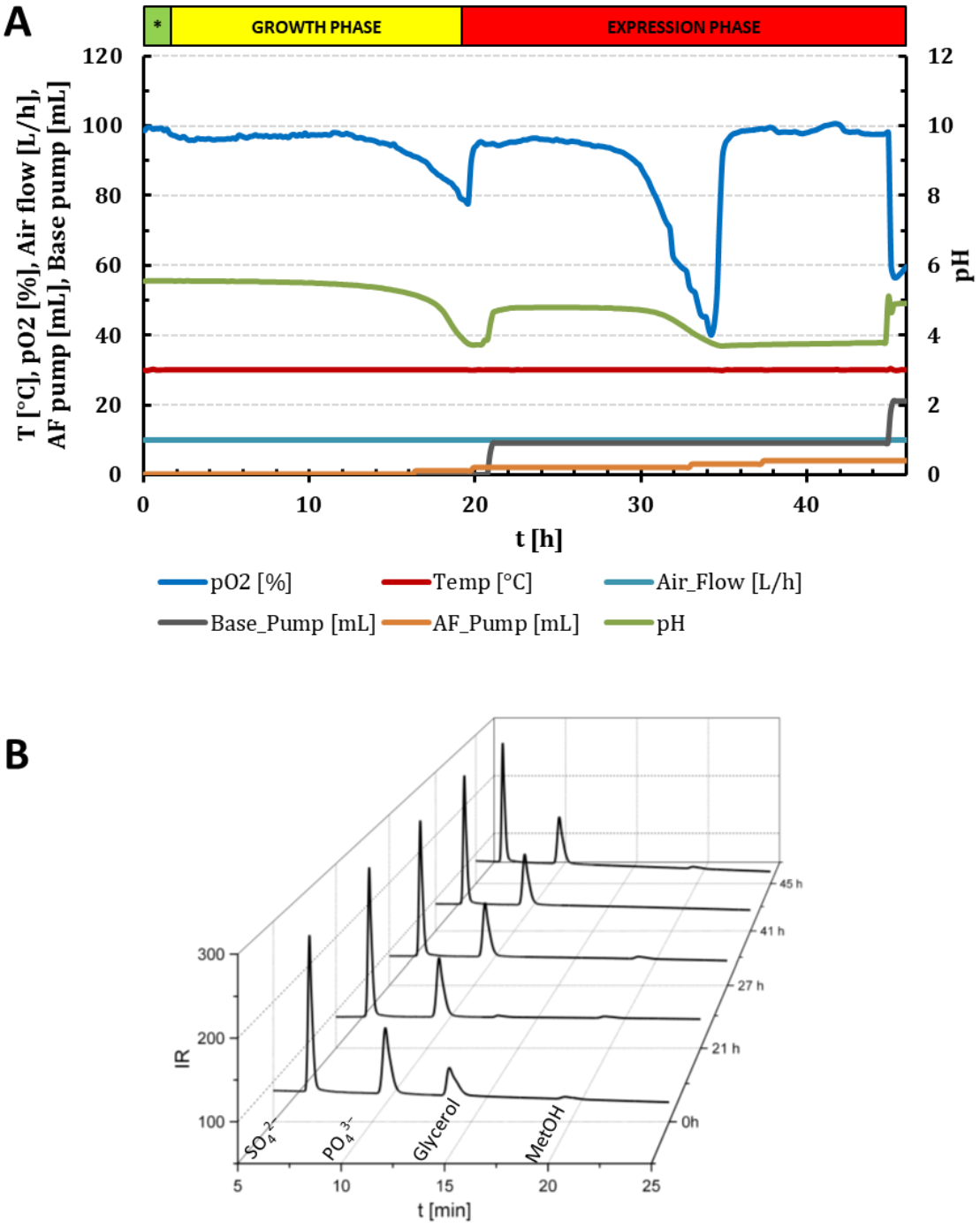


Figure 6: Igni18 production at 10 L scale - fermentation course.
Data represent a typical run of a 10 L fermentation in batch mode. Oxygen partial pressure (pO_2), temperature, pH, air flow and addition of base and antifoam (AF) were monitored by the fermenter (A). *The stability of the fermenter was recorded for 1.5 h prior to inoculation. Samples were taken at the start and after 21, 27, 41 and 45 h. After filter-sterilization, HPLC analysis were performed to quantify both carbon sources, glycerol and MetOH (B).

Centrifugation of the fermentation broth recovered 154 g (wet weight) of yeast cells. After purification by affinity chromatography on Ni-TED agarose, the purity of Igni18_c-Myc_His_6 (from now on referred to as 'Igni18') was estimated to be higher than 90 % (Figure 7) and the yield was approximately 12 mg per liter of expression culture. The purified and recombinant Igni18 has a theoretical MW of 28,614 Da. On an SDS-PAGE, a three-band pattern appears with apparent molecular weights of 30, 60 and 130 kDa. Using a Western Blot immunodetection method, the three bands could be identified as His_6-tagged Igni18 (Figure 7). The native protein seemed to occur in an oligomeric state and this complex is highly stable, as expected from the nature of its native host; it can be denatured only partially with a reducing loading dye and heat incubation for 10 min at 95 °C. Under semi-native running conditions (SDS–PAGE with non-reducing loading dye and without prior heat denaturation) only a single band could be observed on the gel at slightly below 130 kDa (see Figure 16). This suggested that native, soluble Igni18 predominantly exists in a multimeric form.

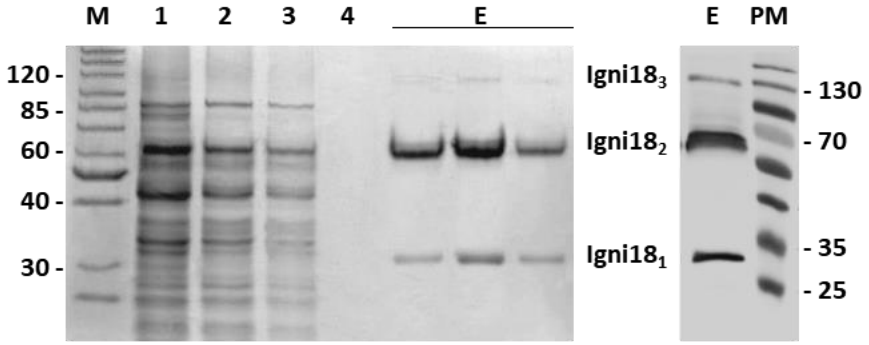


Figure 7: IMAC Purification of Igni18_Myc_His6 (SDS-PAGE).
Left gel was stained with Coomassie Brilliant Blue R-250, right gel was immunodetected (Western Blot) with anti-His-pAB and anti-rabbit-AP antibodies (Carl-Roth, Karlsruhe, Germany). M: protein molecular weight marker (PageRuler Unstained Protein Ladder, Thermo Scientific), 1: lysate, 2: flow-through, 3: initial column wash, 4: final column wash, E: elution fractions, PM: prestained marker (PageRuler Prestained Protein Ladder #26616, Thermo Fisher Scientific). Bands belonging to Igni18 are labelled.

3.1.2 Crystallization and structure determination of Igni18

As of today, only very few examples of crystal structures of archaeal lipases/esterases or MβLs are available in the PDB-database (Table 11). None of the few proteins currently deposited in PDB has been characterized as promiscuous MβL.

Table 11: Crystal structures of archaeal ABHs and MβLs in the PDB-database.
Searches were restricted to entries of archaeal source organisms containing the key words "lipase", "carboxylesterase", "alpha beta hydrolase" and all activities in Table 1 (for MβLs). Presence of the ABH- or MβL-fold was confirmed with a CD-search (Table 10). CPSF: Cleavage and Polyadenylation Specificity Factor; KH: N-terminal K-homology-domains; PTE: phosphotriesterase.

Enzyme class	PDB no.	Cofactor	Source organism
Lipase (ABH)	6QE2	---	*Paleococcus ferrophilus*
Lipase/esterase (ABH)	5LK6	---	*Sulfolobus islandicus*
Lipase (ABH)	2ZYI, 2ZYR	Ca^{2+}/Mg^{2+}	*Archaeoglobus fulgidus*
Lipase (ABH)	2RAU	Ca^{2+}	*Sulfolobus solfataricus*
Esterase (ABH)	3ZWQ	---	*Pyrobaculum calidifontis*
Carboxylesterase (ABH)	5A62	---	*Nitrososphaera gargenis*
Carboxylesterase (ABH)	3AIK	---	*Sulfolobus tokodaii*
Carboxylesterase (ABH)	2C7B	---	Uncultured Archaeon
Carboxylesterase (ABH)	1JJI	---	*Archaeoglobus fulgidus*
MβL	3ADR	Zn^{2+}	*Sulfolobus tokodaii*
CPSF (MβL)	3AF5	Zn^{2+}	*Pyrococcus horikoshii*
KH-CPSF (MβL)	2XR1	Zn^{2+}	*Methanosarcina mazei*
KH-CPSF (MβL)	2YCB	Zn^{2+}	*Methanothermobacter thermautotrophicus*
RNase J (MβL)	5HAA	Zn^{2+}	*Methanolobus psychrophilus*
Lactonase/PTE (MβL)	2VC5	Fe^{3+} + Co^{2+}	*Saccharolobus solfataricus*

Within this framework, the 3D structure of Igni18 was established. The protein crystallized in the space group R32 and crystals diffracted to a resolution of 1.7 Å (Figure 8 and Supplementary Table 2). Data were collected at beamline ID30A-3 (ESRF, Grenoble, France) and processed as already described (Kobus *et al.* 2019).

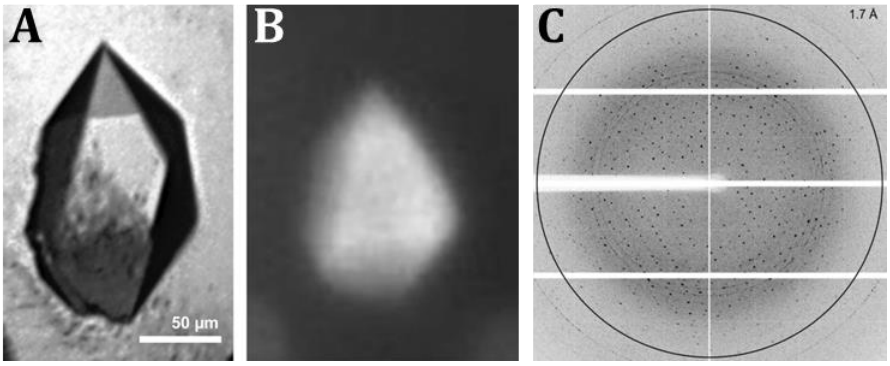


Figure 8: Crystallization and X-ray diffraction of Igni18.
Visible image (A) and UV image (B) of an Igni18 crystal obtained in 0.3 M magnesium nitrate hexahydrate, 0.1 M Tris pH 8, 22 %(w/v) PEG 8000. Diffraction image (C) with a resolution of 1.7 Å at the edge of the detector (indicated by the black circle). Figure adapted from Kobus *et al.* (2019).

The structure was solved using a single SAD dataset to a resolution of 2.3 Å. The protein model was then further built and refined to R values of 18.9 % (R_{work}) and 25.7 % (R_{free}) with 93.5 % of the residues being in the favored region of the Ramachandran plot (see Supplementary Table 3 for data collection and refinement statistics). The Igni18 coordinates were deposited in the PDB under the accession code 6HRG. Although the ASU contains only one monomer, the biologically active arrangement of Igni18 is a trimer, as is also found in the crystal lattice via its symmetry-related molecules (Figure 9A). Igni18 is the first (and to date only) archaeal crystal structure available in the PDB to be produced heterologously in *P. pastoris* (or *Komagataella phaffii*). Together with the ornithine carbamoyltransferase from *Pyrococcus furiosus* (1A1S and 1PVV) which was expressed in *Saccharomyces cerevisiae* (Villeret *et al.* 1998, Massant *et al.* 2003), they are the only archaeal crystal structures available that were produced using the two most common yeast expression systems.

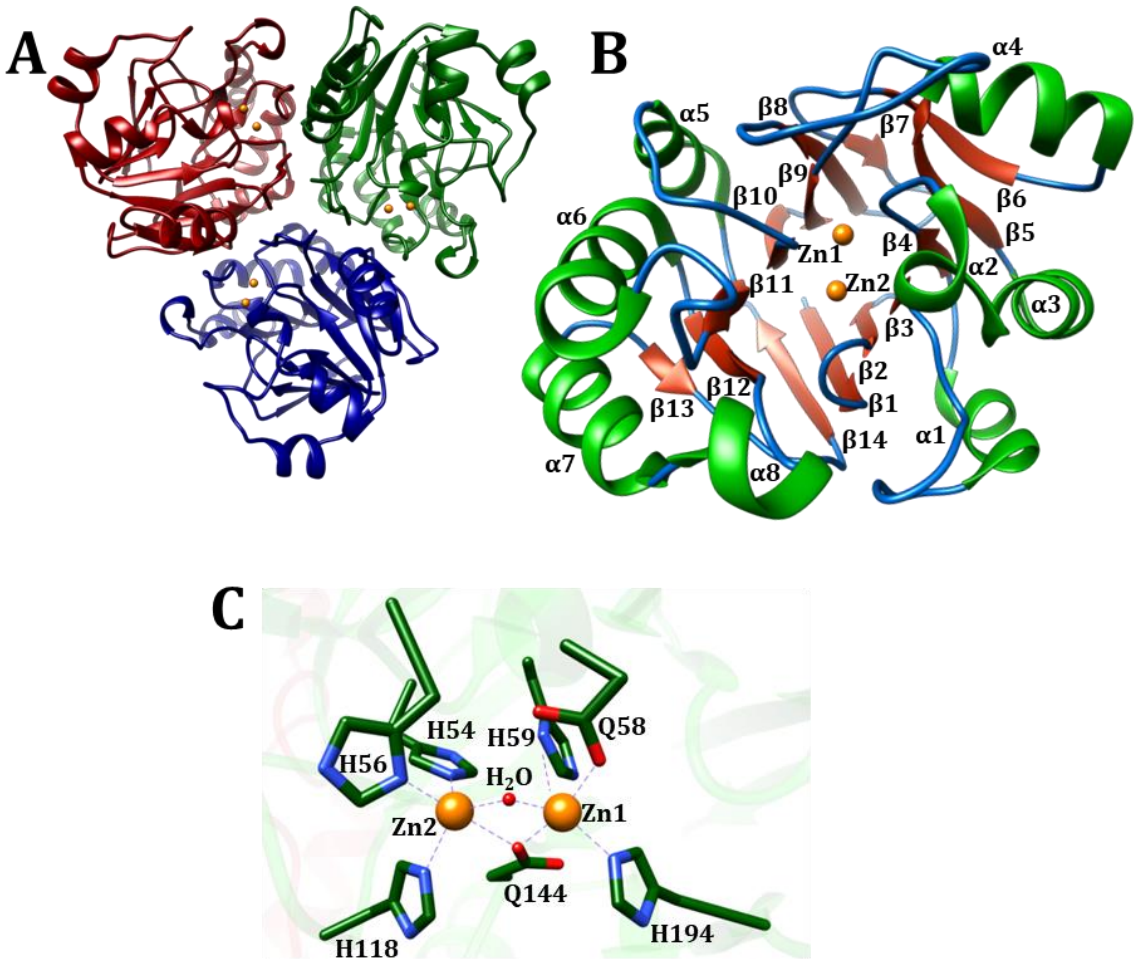


Figure 9: Crystal structure of Igni18.
The protein crystalized as a homotrimer (A). A monomer (B) comprises 8 α-helixes (green) and 14 β-strands (red) and contains two metal ions. A close-up of the catalytic pocket at the interface between two monomers shows the metal coordination (C). Zn^{2+} ions are depicted in orange.

The overall fold of each monomer is composed of two mixed β-sheets (β14 plus β1 to β6 and β7 to β13) slightly twisted and arranged in a parallel manner in the center of the protein and flanked by four helixes on each side of the sheets (α1 to α4 and α5 to α8; Figures 2A and 9B), thereby building the four-layered αββα core typical of the MβL superfamily (Carfi *et al.* 1995, Li de la Sierra-Gallay *et al.* 2005). The interfaces between the three monomers are all ~850 Å^2 with 11 intermolecular hydrogen bonds: His78-'Tyr121,

Glu79-‘Ser119, Glu79-‘His56, Glu79-‘Tyr121, Asn95-‘His170, Asn95-‘Ser119, Gly97-‘Gln177, Gly98-‘Thr172, Gly98-‘His170, Glu150-‘Ala149 and Tyr158-‘Gln177. Close below the protein surface and in direct proximity to the loops connecting β4-α2, β10-α5 and β11-α6, two metal ions are bound by His54, His56, His59, His118, His194, Asp58 and Asp144 (Figures 9B and C).

ICP-MS analysis conducted to verify the chelated ions within the catalytic pocket revealed Zn being preferably bound to Igni18 under recombinant *in vivo* conditions (Figures 9A and B), with 87.5 % mass fraction (Figure 10). Ni, Fe and Mn accounted for 6.8, 3.9 and 1.5 % (w/w) respectively. Cu (0.4 % w/w) and Co were barely present or not at all. Other metals (Ti, Se, Nb, Mo, Ag, Cd, Hg and Pb) represented only 0.45 % (w/w) of the total detected mass.

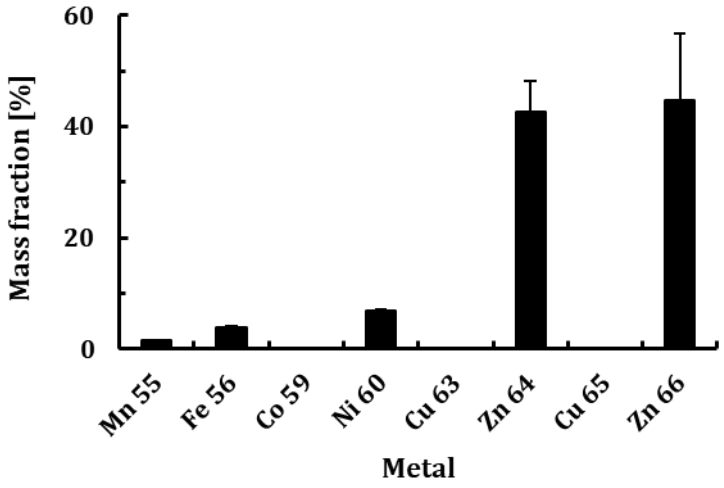


Figure 10: ICP-MS analysis of in vivo produced and purified Igni18.
Zn was the metal preferably bound by Igni18. Ni, Fe and Mn were also detected. The mean values of three measurements are represented. Bars indicate relative standard deviation (RSD).

3.1.3 Biochemical characterization

Prior to activity assays, the obtained crystal structure was uploaded to the COFACTOR server to identify possible functional sites and homologies. The best Gene Ontology hits are shown in Table 12.

Table 12: Structure-based molecular function prediction.
The annotation was performed with the COFACTOR server (Zhang *et al.* 2017). Local and global structure matches against the BioLiP protein function database were executed to identify functional sites and homologies. Gene Ontologies (GOs) and their scores were derived from the best functional homology templates.

GO term	CscoreGO	Name of predicted function
GO:0004620	1.00	Phospholipase activity
GO:0008081	0.91	Phosphoric diester hydrolase activity
GO:1901363	0.89	Heterocyclic compound binding
GO:0097159	0.89	Organic cyclic compound binding
GO:0070290	0.86	Phospholipase D activity
GO:0000166	0.69	Nucleotide binding
GO:0005524	0.57	ATP binding

MβLs require the coordination of at least one metal ion for displaying activity. For the following characterization, purified Igni18 was treated with the chelating agent EDTA to remove ions bound during the *in vivo* production or purification by affinity chromatography (Ni-TED). After the treatment, stripped Igni18 could be supplied with different metals at will.

Hydrolysis of β-lactam antibiotics: Class B MβLs name the protein fold of Igni18, but β-lactams are generally heat labile and the organisms producing them grow mostly at mesophilic temperatures. To test if Igni18 displays β-lactamase activity, disk susceptibility assays were performed. Analyses were not conducted at 90 °C, which is the optimal growth temperature of *I. hospitalis*, but instead at 40 °C overnight. Igni18 supplied with Ni^{2+} or Zn^{2+}

as cofactor showed significant β-lactamase activity against most of the β-lactam antibiotics tested (Figure 11). Thereby Igni18 activity was metal-dependent. In the presence of Zn^{2+}, the penam mezlocillin (MEZ) and the carbapenem imipenem (IPM) were fully inactivated. With Ni^{2+}, 75 % reduction of the zone of inhibition (ZOI) of MEZ and 100 % of IPM was observed. Cephalosporines from the second and third generation were partially degraded by Igni18. The ZOI of cefamandole (MA) and cefotaxime (CTX) were reduced by 88 % and 82 % with Zn^{2+} as a cofactor; loracarbef (LOR) and cefaclor (CEC) by 48 % and 42 %. There was no observable activity on cefotiam (CFT). With Ni^{2+} as a cofactor, the highest activity was observed against CTX (16 % reduction). The ZOI of CFT, MA, LOR and CEC were reduced by less than 10 % or not at all.

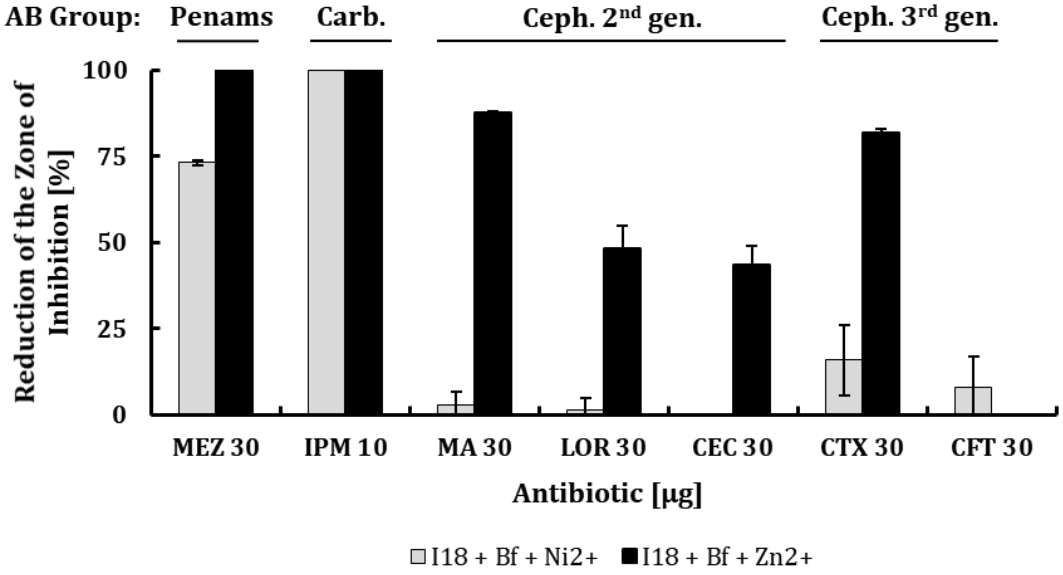


Figure 11: The β-lactamase activity of Igni18 is highly metal-dependent.
Antibiotic discs were incubated with and without Ni^{2+}- and Zn^{2+}-containing buffer together with Igni18 over night at 40 °C. The clearing zones of the discs on agar plates with *E. coli* DH5α were measured in triplicate after another night of incubation at 37 °C. Based on the halo´s diameter, the reduction of the zone of inhibition [mm^2] by Igni18 was calculated. Mezlocillin 30 µg (MEZ 30), imipenem 10 µg (IPM 10), cefamandole 30 µg (MA 30), loracarbef 30 µg (LOR 30), cefaclor 30 µg (CEC 30), cefotaxime 30 µg (CTX 30) and cefotiam 30 µg (CFT 30). Carb.: Carbapenemes; Ceph.: Cephalosporines. Data are the mean values of 3 measurements and bars represent standard error.

Esterase activity profile of Igni18 with *p*NP-substituted model substrates: The esterase activities of Igni18 in dependency on metal ions were assayed on various esters containing the *p*NP chromophore. Carboxyl-esterase and lipase activities (EC 3.1.1.1) were tested on *p*NP esters with acyl chains containing 2 to 18 carbon atoms, phosphatase activity (EC 3.1.3) with *p*NPP, phosphodiesterase activity (PDE, EC 3.1.4) on bis-*p*NPP and *p*NPPP, phosphotriesterase activity (PTE, EC 3.1.8.1) on paraoxon and parathion, and phospholipase C activity (PLC, EC 3.1.4.3) on *p*NPPC. The pH and temperature optima of Igni18 were determined using a representative substrate for each enzyme family (*p*NP-C16, bis-*p*NPP, paraoxon and *p*NP-PC; Figure 12).

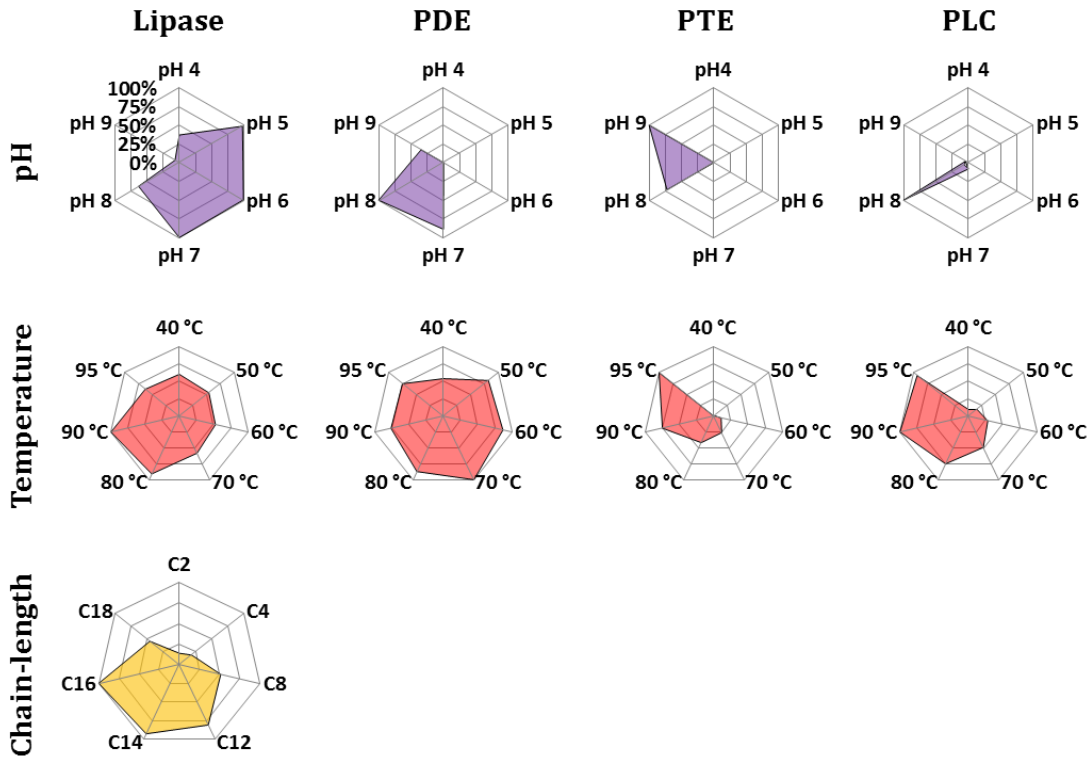


Figure 12: Determination of optimal pH and temperature for diverse activities of Igni18.
Relative activities are depicted. Lipase activity, temperature and pH optima were determined on *p*NP-palmitate (*p*NP-C16), phosphodiesterase activity (PDE) on bis-*p*NPP, phosphotriesterase activity (PTE) on paraoxon and phospholipase C activity (PLC) on *p*NP-PC. Data are the mean values of three measurements; standard error ranged from 1 % to 10 %.

Remarkably, lipase activity was measurable between pH 5 and 8, while the other activities were only detectable in narrower pH ranges above 7. No remarkable activities were measured at pH 10 (data not shown). Lipase and PDE activities were measured at a broad temperature range between 40 °C and 95 °C, while PTE and PLC activities were only measured above 60 °C.

While hydrolysis of carboxyl-esters was not strictly depending on the presence of metal ions, it could be that the metal bound to Igni18 strongly influences the other activities assayed (Figure 13). Carboxyl-esterase activity could be measured on all substrates assayed, but long chain fatty acid esters (C12-C16) were preferred (Figure 12). The highest activity was measured with the typical lipase substrate *p*NP-palmitate. Surprisingly, the presence of Mn^{2+}, Cu^{2+} or Zn^{2+} raised carboxyl-esterase activity by 10 to 20 % and Ni^{2+} lowered it by almost 50 % (Figure 13). The enzyme showed esterase activity in the pH range of 4 to 8 (optimum between pH 5 to 7) and over 50 % activity was observed over the whole temperature range tested, with its maximum at 90 °C (Figure 12). Phosphatase activity was not observed with *p*NPP as the substrate, while PDE activity was observed on both bis-*p*NPP and *p*NPPP and was strictly dependent on the presence of Ni^{2+}. The presence of Mn^{2+}, Fe^{2+} or Co^{2+} resulted in activities lower than 10 % of the activity observed with the main ion. With bis-*p*NPP and *p*NPPP, the highest activity was measured at pH 8. Less than 50 % activity was retained at pH 9 and no hydrolysis was found at values below 7. Hydrolysis of bis-*p*NPP was remarkable (more than 75 % of the highest activity) at temperatures above 50 °C, with the highest activity determined at 70 °C. The hydrolysis of paraoxon was strictly Zn^{2+}-dependent. Hydrolysis of parathion, although occurring primarily in the presence of Zn^{2+}, was also observed in the presence of Co^{2+} (60 % activity)

and Mg^{2+}, Fe^{2+} and Ni^{2+} at rates below 25 % of the maximum activity. The highest activity was observed at pH 8 to 9 and temperatures between 90 and 95 °C. Finally, PLC activity was found to depend on Ni^{2+} as a cofactor (Mn^{2+} yielded less than 10 %) in the temperature range from 70 to 95 °C (optimum at 90 °C) and only at pH 8.

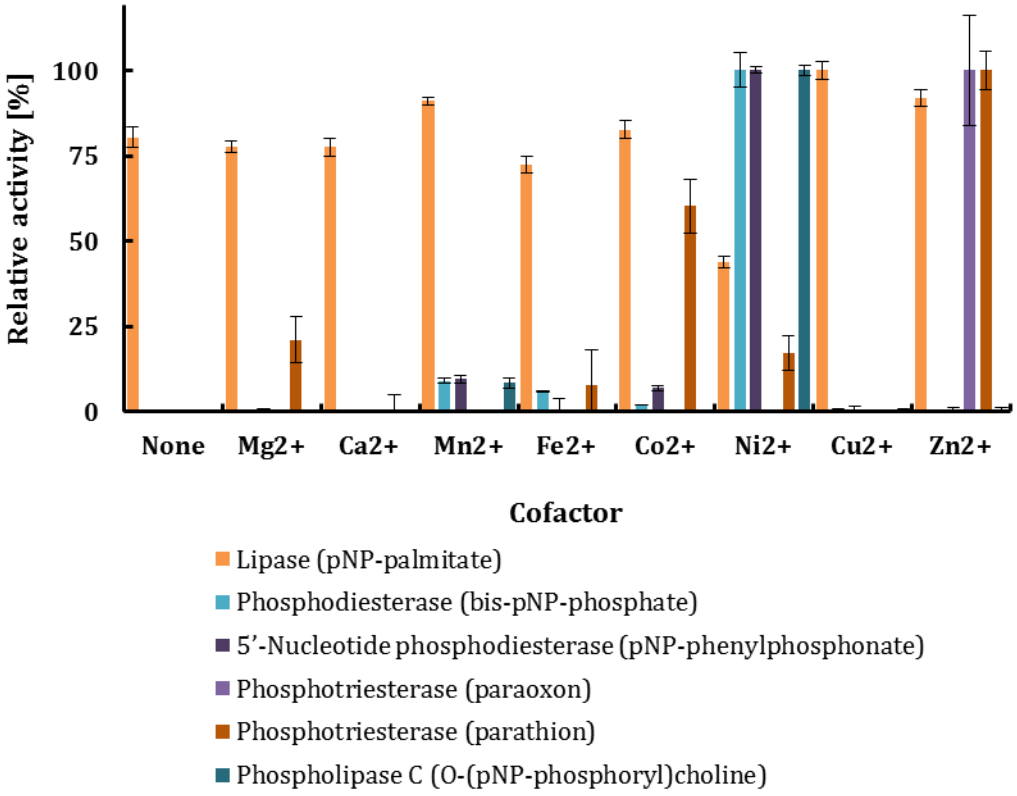


Figure 13: Influence of metal-ion cofactors on Igni18´s promiscuous activities.
Activities were assayed at 90 °C. Data represent mean values of triplicates and bars indicate standard error.

In order to find out which one of these substrates was hydrolyzed at the highest rate, the enzymatic units were all measured at 90 °C under optimal pH and metal conditions. The highest activities were registered on the PDE substrates bis-*p*NPP (1.26 U/mg) and *p*NPPP (1.72 U/mg). Igni18 showed 0.45 U/mg PLC activity, and only 0.08, 0.03 and 0.004 U/mg were measured on paraoxon, parathion and the lipase substrate *p*NP-C16, respectively

(Figure 14). Michaelis-Menten kinetic parameters determined with the bis-*p*NPP substrate at 90 °C and pH 8 were v_{max} = 7.04E-10 mol/min; K_M = 0.0025 mM and k_{cat} = 6.72E-5 s^{-1} (Supplementary Figure 4).

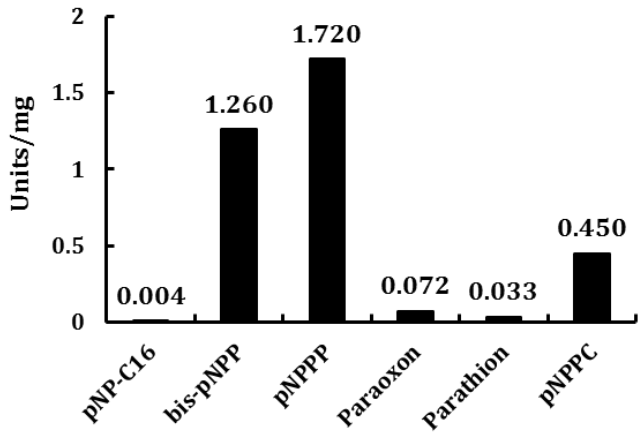


Figure 14: Activity of Igni18 on various substrates (in U/mg).
Activities were assayed at 90 °C with 1 mg Igni18 per mL stock solution for all activities but on bis-*p*NPP and *p*NPPP (0.1 mg/mL). Data represent mean values of three replicates.

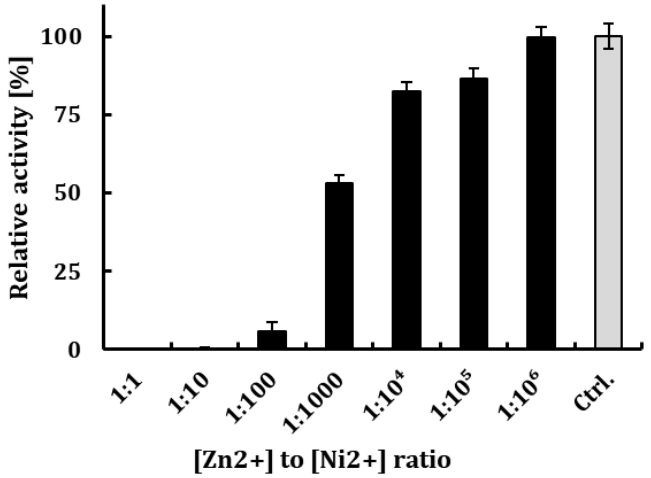


Figure 15: Ion competition assay between Zn^{2+} and Ni^{2+}.
Both metals were premixed in the mentioned ratios prior to addition to the reaction mixture. A Ni^{2+}-only control was included. Activity was assayed at 90 °C on bis-*p*NPP. Data represent mean values of triplicates and bars indicate standard error.

Both Zn^{2+} and Ni^{2+} seem to get incorporated into Igni18 resulting in an active enzyme, but the second seems to bind stronger (Figures 10, 11 and 13). To determine which one is preferably bound within the catalytic pocket, metal-free Igni18 was supplemented with different ratios of both metals and activity was assayed on bis-*p*NPP (Figure 15). No activity could be detected at Zn:Ni ratios 1:1 or 1:10. At 1:100 ratio, only 6 % of Igni18 was active (loaded with Ni^{2+} ions) and 1:1,000 yielded 53 % activity. Ratios of $1:10^4$ and $1:10^5$ revealed more than 80 % activity, but full activity compared to the control could only be recovered if Ni^{2+} was present in 10^6-fold molar excess to Zn^{2+}.

3.1.4 Thermal stability, unfolding and refolding

Igni18 is an extremely thermostable enzyme indicated by a half-life time of approx. 48 h at 90 °C (Pérez-García *et al.*, submitted). Preliminary experiments showed no loss of quaternary structure or activity after subjecting the protein to several freeze-thaw cycles or incubating it at 99 °C for 10 min, as the trimer band was observable on a semi-native PAGE (Figure 16). Partial degradation could only be achieved in combination with SDS.

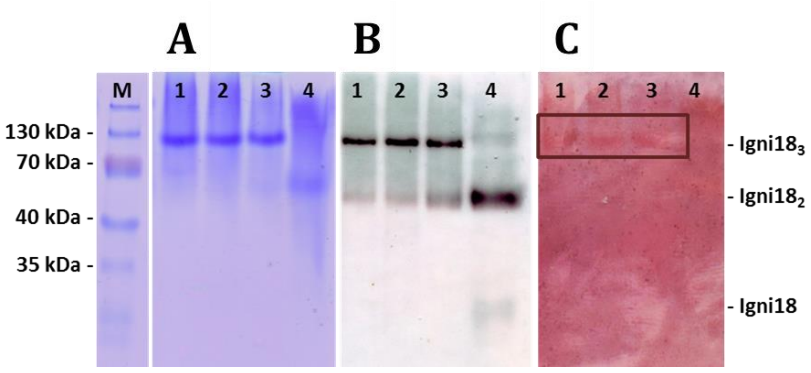


Figure 16: Igni18 is highly thermostable and biologically active in its trimeric form.
12 % acrylamide gels were loaded with native, non-denatured Igni18 protein samples produced in *P. pastoris*. Protein bands were visualized by Coomassie staining (A), Western Blot immunoassay using Histidine-specific antibodies targeting the tagged protein (B) and a zymogram analysis with α-naphthyl acetate and Fast Blue reagent (C). M, Pre-stained marker; 1, purified Igni18, native, untreated; 2, purified Igni18, native, after three freeze-thaw cycles; 3, purified Igni18, native, incubated for 10 min at 99 °C; 4, purified Igni18 denatured with SDS at 95 °C for 10 min.

To test the thermal stability of Igni18, unfolding was measured by the changes of intrinsic protein fluorescence upon heating with the nanoDSF method. These experiments were performed by collaboration partners at the Heinrich Heine Universität Düsseldorf (Christoph Strunk, Filip Kovacic and Karl-Erich Jäger). Igni18 contains a single aromatic residue (W28) partially exposed on the protein surface, but distant from the interface between the monomers, representing a suitable system for such an analysis (Figure 17).

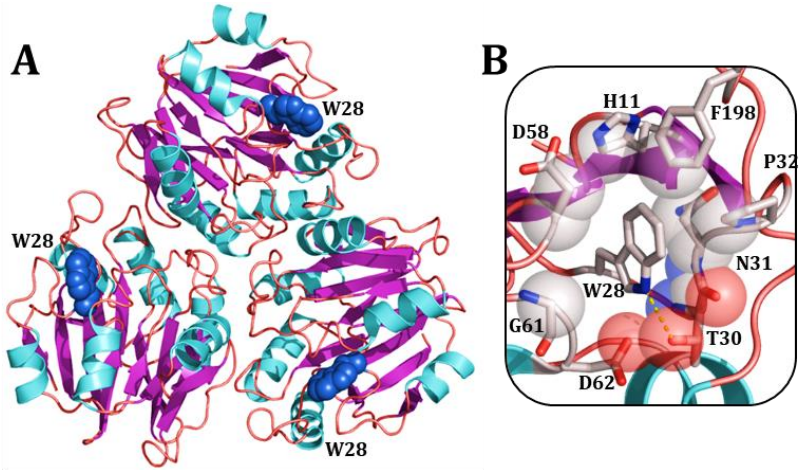


Figure 17: Localization of the Trp residues used for nanoDSF experiments.
The aromatic residue is located partially exposed on the protein surface, but distant from the interface between the monomers (A). The binding pocket of W28 shows a strong hydrophobic character (B). Adapted from Pérez-García *et al.* (submitted).

The thermal unfolding curve of Igni18 recorded between 15 °C and 110 °C shows two transitions, one in the range of 55-65 °C and the second one at approx. 95 °C (Figure 18A). The melting temperature for the low-temperature unfolding increased as the heating rates increased while the high-temperature unfolding transition did not show a heating rate dependency (Figure 18A). The nanoDSF results indicate that the low-temperature unfolding is a kinetically controlled process representing a hallmark of irreversible aggregation (Pauwels *et al.* 2012).

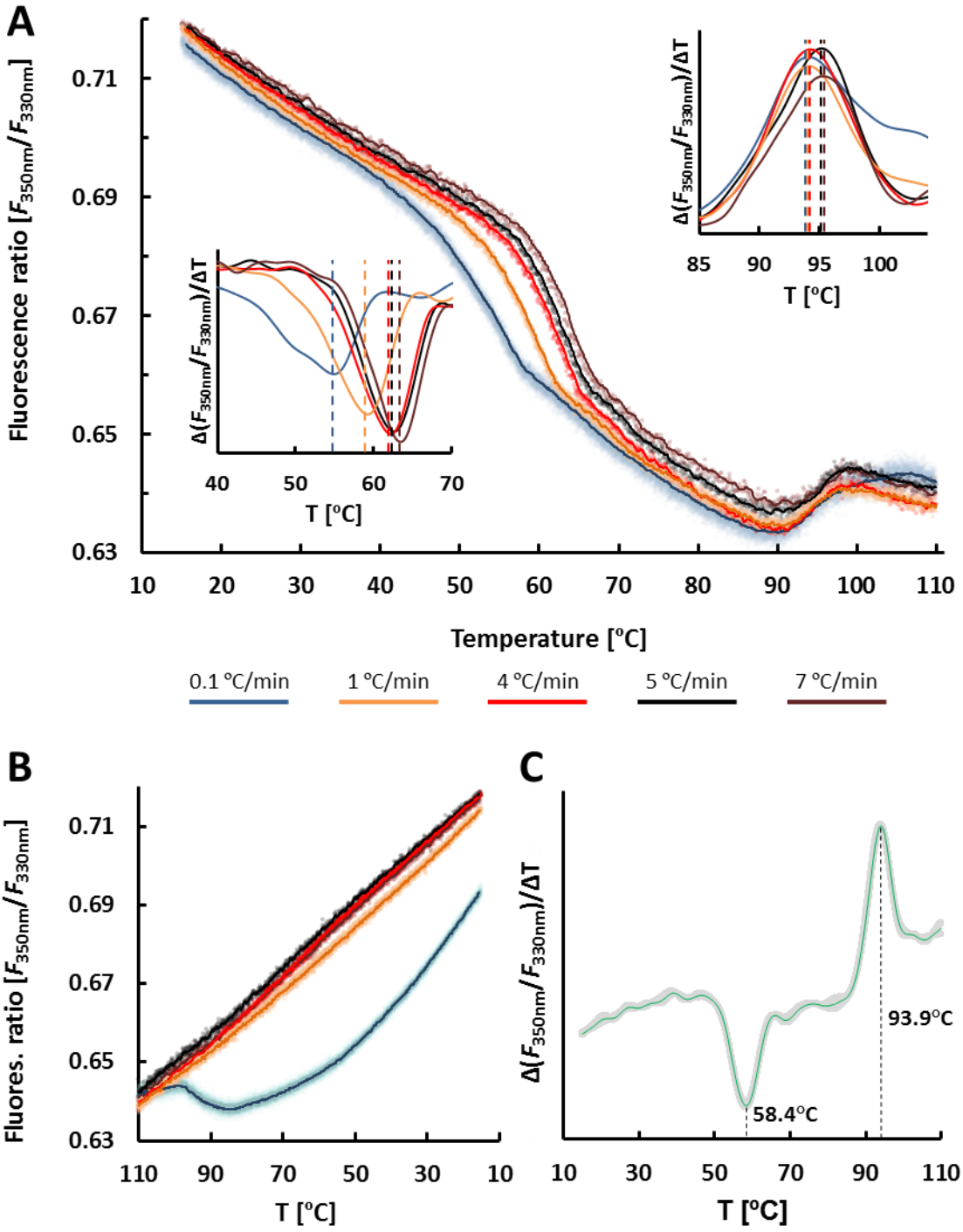


Figure 18: Two-step thermal unfolding pathway of Igni18.
Effect of temperature heating at different rates on the thermal unfolding of native Igni18 analyzed by nanoDSF, measuring the intrinsic protein fluorescence from 15 °C to 110 °C (A). Thermal refolding of Igni18 (B). Thermal unfolding of Igni18 in the presence of EDTA shows that metal cations do not have a heat-stabilizing role (C). Adapted from Pérez-García *et al.* (submitted).

However, reversible unfolding of Igni18 was observed. The protein, which has been heated to 70 °C and then cooled, showed not only complete restoration of phosphodiesterase activity, but even a 50 % increase (Figure 19). In contrast, the irreversibility of high-temperature unfolding was proposed from phosphodiesterase activity measurements (Figure 19).

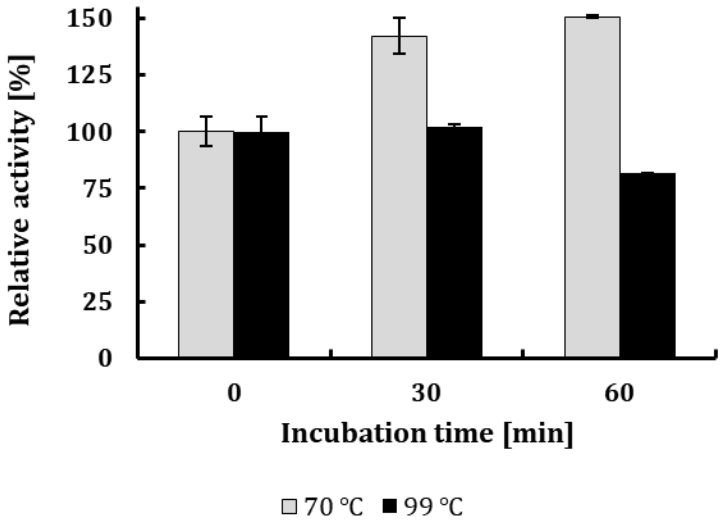


Figure 19: Reversible and irreversible unfolding of Igni18.
Protein was unfolded for 30 and 60 min at both temperatures, refolded at room temperature for 15 min and phosphodiesterase activity was assayed on bis-*p*NPP.

To test if high temperature unfolding leads to irreversible aggregation, nanoDSF refolding experiments were performed in which native Igni18 was first heated up to 110 °C followed by cooling to 20 °C with identical heating and cooling rates. The results exposed that Igni18 regained fluorescence after refolding to reach a final fluorescence almost identical to the initial fluorescence of the non-heated protein (Figure 18B). Heating of Igni18 to 110 °C did not lead to irreversible aggregation, as a refolded form of Igni18 showed comparable fluorescence properties as the native form. Overall, these unfolding/refolding tests and

enzyme activity measurements indicate a two-step unfolding pathway for Igni18. Native Igni18 is partially unfolded at ~55-65 °C and can be reversibly converted to a fully active enzyme by cooling. Further heating of Igni18 above 95 °C eventually results in a loss of enzymatic activity (Figure 19), which cannot be refolded by cooling, but is not prone to aggregation (Figure 18).

The activity assays and the X-ray structure of Igni18 showed that divalent cations have an important role for the enzymatic function (Figures 9, 10, 11 and 13). To find out if these divalent metal cations also have a structural role, Igni18 was incubated with EDTA in order to remove the cations bound to the protein before thermal unfolding measurements were conducted. The melting temperatures of metal-free Igni18 of 58.4 °C and 93.9 °C were nearly equal to those of native Igni18 (Figure 18C). This indicates that metal cations do not stabilize the structure of Igni18.

3.1.5 Sequence-based searches against NCBI's non-redundant database

A BLASTP search against the nr-database revealed homologs of Igni18 belonging to different lineages as shown by the molecular relations among them (Figure 20). Within the first 1,000 homologs (score ≥ 150; e-value ≥ 6E-42; sequence identity ≥ 35.9 %), proteins were mostly annotated as metal- or Zn-dependent hydrolases from both Archaea and Bacteria. Within the Archaea, most of the hits belonged to both Crenarchaeota and Euryarchaeota, as well as to the groups Aigarchaeota, Bathyarchaeota, Korarchaeota and Geo(therm)archaeota. Full-length hits for Thaumarchaeota or Archaea from the Asgard group (excluding Thorarchaeota) were not found in this reduced dataset but do exist with lower scores (≤ 86.3). Most of the bacterial homologs were found to belong to the groups Deltaproteobacteria and Firmicutes, but also in a lower share to Actinobacteria, green

sulfur and non-sulfur Bacteria, Planctomycetes, α- and β-Proteobacteria, Verrucomicrobia and Aquificales amongst others. The best homolog belonging to the Eukarya domain is the hypothetical protein SARC_07738 from *Sphaeroforma arctica* JP610 (Score 144, e-value 2E-39, seq. id. 37.3 %). Other hits of eukaryotic origin include putative metallo-hydrolases/oxidoreductases, glyoxalase II, UlaG or NAPE-PLD at lower scores (Table 1).

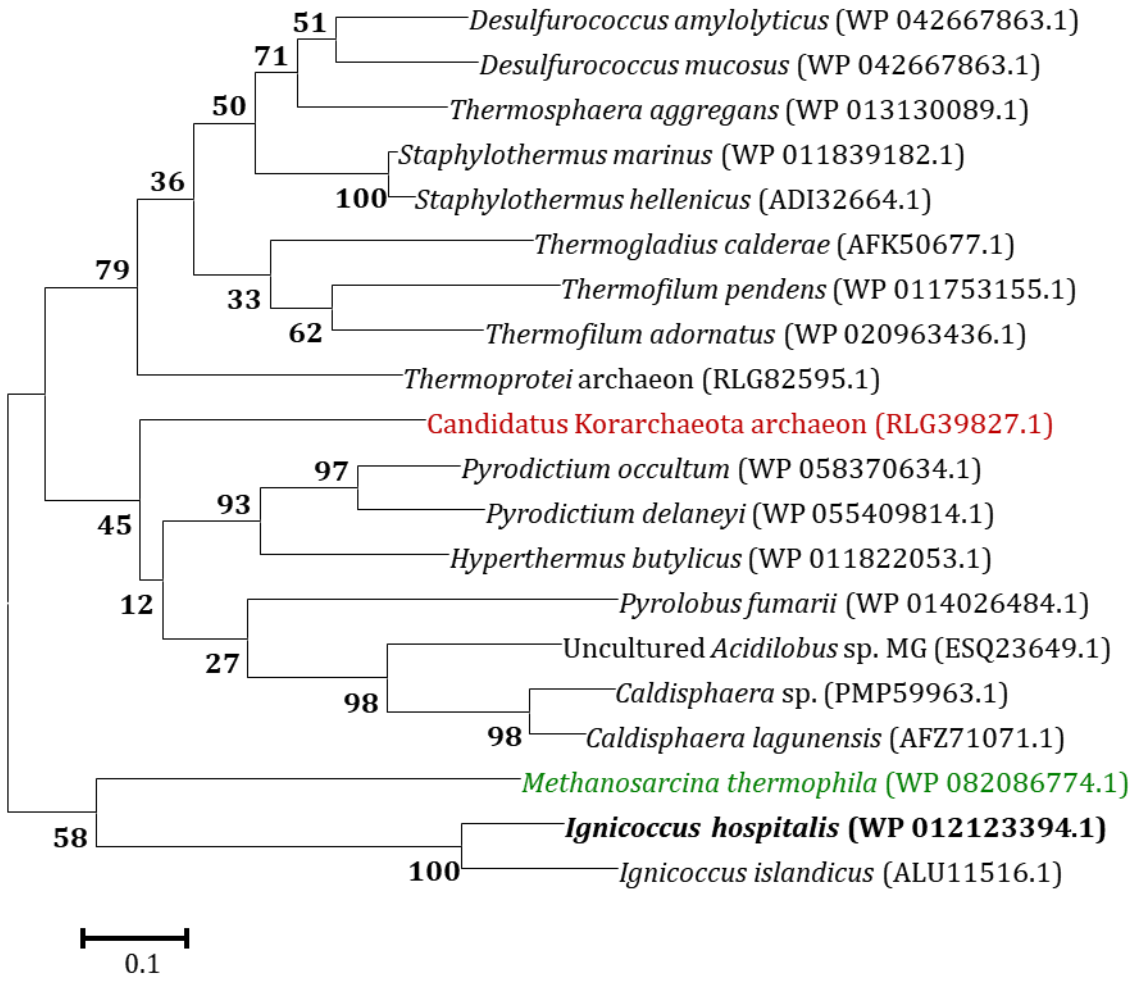


Figure 20: Protein sequence relations between Igni18 and its homologs.
Crenarchaeal sequences are depicted black (Igni18 is shown bold), Korarchaeota in red and Euryarchaeota in green.

Protein sequence comparison of Igni18 with its nine closest homologs revealed high amino acid conservation within different archaeal species. The five His and two Asp needed for metal coordination are present in all of them. Moreover, numerous motifs with sequence identities of more than 80 % were found within the alignment (Figure 21).

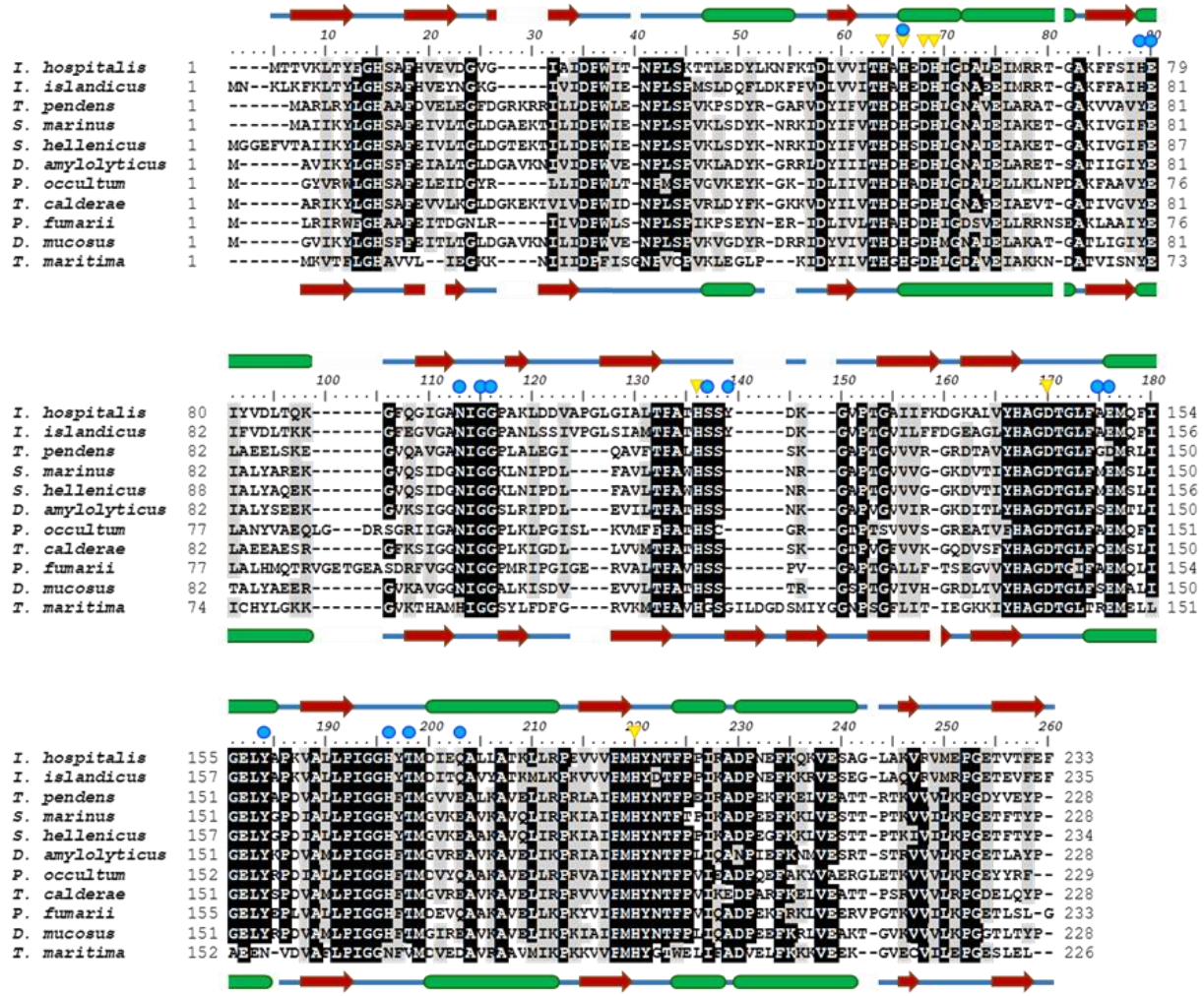


Figure 21: Sequence comparison between Igni18 and its homologs.
Nine homologs were selected for sequence comparison. Each sequence is labeled by the gene identification number: WP_012123394.1 , Igni18 (*I. hospitalis*); WP_075049418.1 (*Ignicoccus islandicus*); WP_011753155.1 (*Thermofilum pendens*); WP_011839182.1 (*Staphylothermus marinus*); WP_052833686.1 (*Staphylothermus hellenicus*); WP_042667863.1 (*Desulfurococcus amylolyticus*); WP_058370634.1 (*Pyrodictium occultum*); WP_048162871.1 (*Thermogladius calderae*); WP_014026484.1 (*Pyrolobus fumarii*); WP_013562750.1 (*Desulfurococcus mucosus*). WP_004080208.1 (*Thermotoga maritima*) is also included in the analysis. Black shading represents conservation of more than 80 %, grey shading highlights similar amino acid residues. The secondary structures derived from Igni18 and the metallo-hydrolase from *T. maritima* are shown respectively above and below its sequence (α-helixes in green, β-sheets in red). The seven amino-acid residues necessary for metal coordination are marked with a yellow inverted triangle (∇); the ones needed for trimerization are highlighted with a blue circle (○). Figure from Pérez-García *et al.* (submitted).

3.1.6 Structural alignments against the Protein Data Bank

Structural comparison of an Igni18 monomer against the PDB revealed hits belonging to very diverse protein families, with enzymes annotated as NAPE-PLD, UlaG, RNase J, ICTase, diiron β-hydroxylase, CPSF1 or RNase Z, with total RMSD values of 2 onwards (Table 1, Figure 22 and Supplementary Table 4). These hits display a very similar core structure, but some loops and flanking regions showed high diversity. The best hit belonged to the crystal structure of a MβL from the thermophilic bacterium *Thermotoga maritima* (Choi *et al.* 2015) with a RMSD-value of 0.82 over 173 Cα atoms for the monomer. Among many similarities, the *T. maritima* MβL has a MβL core, forms a trimer and showed in addition to its β-lactamase activity also phosphodiesterase activity (Choi *et al.* 2015).

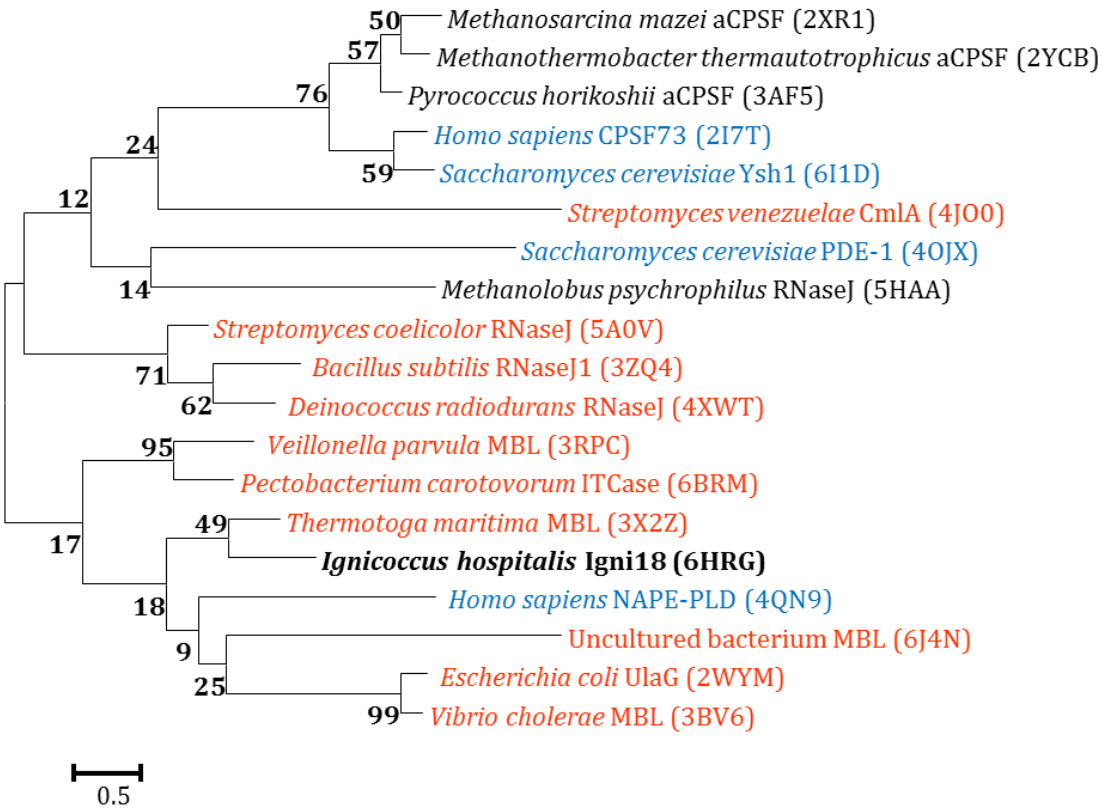


Figure 22: Protein sequence relations between Igni18 and its structural homologs.
Archaeal enzymes are displayed black (Igni18 is bold), bacterial in orange and eukaryotic in blue.

3.1.7 Structural evolution of MβLs

In order to define and measure the diversity and evolution within the MβL family, a subset of crystal structures was selected, and a structural alignment was performed. It included the structures of the MβL from *T. maritima* [3X2Z (Choi *et al.* 2015)], human NAPE-PLD [4QN9 (Magotti *et al.* 2015)], UlaG from *E. coli* [2WYM (Garces *et al.* 2010)], a putative MβL from *V. cholerae* (3BV6), ITCase from *Pectobacterium carotovorum* [6BRM (van den Bosch *et al.* 2018)], RNase J from *Methanolobus psychrophilus* [5HAA (Zheng *et al.* 2017)], β-hydroxylase CmlA from *Streptomyces venezuelae* [4JO0 (Makris *et al.* 2013)], CPSF1 from *Methanosarcina mazei* [2XR1 (Mir-Montazeri *et al.* 2011)] and RNase Z from *Bacillus subtilis* [4GCW (Pellegrini *et al.* 2012); see Figures 23 and 24 and Supplementary Figure 5].

The active pockets and amino acids needed for metal coordination seem to be very conserved (Figure 24). As a consensus, two Zn^{2+} ions are chelated between five His and two Asp. Nevertheless, some of the examined proteins chelate two Ni^{2+} (3X2Z), two Fe^{2+} (3BV6 and 4JO0) or one Mn^{2+} (2WYM). No alterations in amino acid arrangement were found in the MβL from *T. maritima* (3X2Z) and all Zn-binding enzymes. The second coordination sphere (His8 and Glu73') is believed to contribute to the specificity of Ni-binding (Choi *et al.* 2015). Igni18, binding Zn preferably under laboratory conditions, also has these amino acids in the homologous 11 and 79' sites. Fe-binding proteins do not contain the fourth His (His118 from Igni18, Figure 9C). Instead, they possess Asp184 (3BV6) and Glu377 (4JO0; Figure 24). Additionally, the β-hydroxylase 4JO0 has an Ala482 instead of Igni18's His194. UlaG (2WYM), being a very similar protein to *V. cholerae´*s 3BV6, also contains an Asp184, but binds a single Mn^{2+} instead.

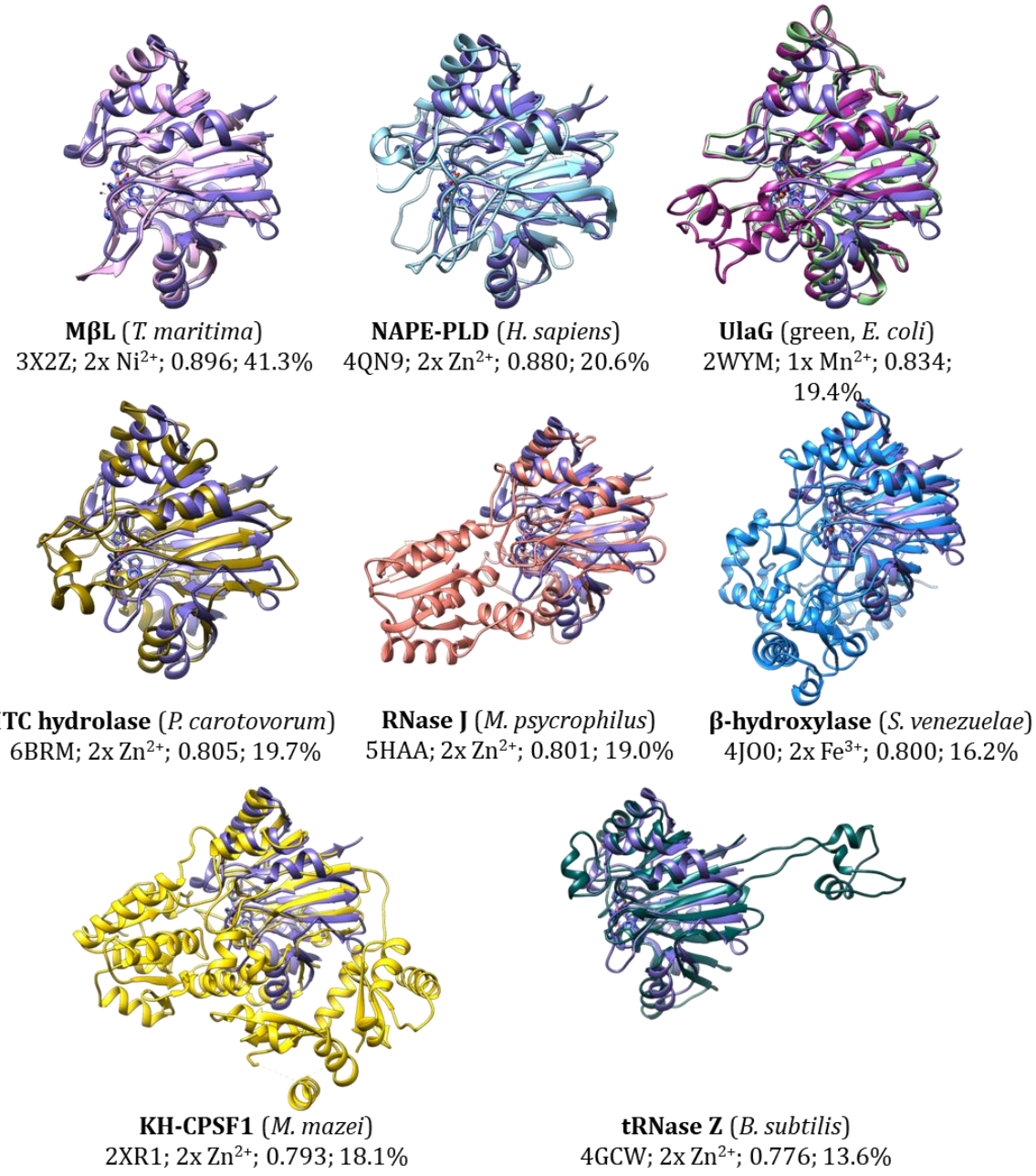


Figure 23: Structural conservation of the Igni18-like core-domain.
The homolog structures 3X2Z from *Thermotoga maritima*, 4QN9 from *Homo sapiens*, 3BV6 from *Vibrio cholerae*, 2WYM from *Escherichia coli*, 6BRM from *Pectobacterium carotovorum*, 5HAA from *Methanolobus psycrophilus*, 4JO0 from *Streptomyces venezuelae*, 2XR1 from *Methanosarcina mazei*, and 4GCW from *Bacillus subtilis* are depicted. Igni18 is depicted in purple. Enzymatic function, PDB code, metal ions bound, structural identity score (TM-score) and sequence identity [%] are given. The uncharacterized 3BV6 protein from *V. cholerae* (2x Fe^{3+}; TM-score 0.8418; seq. id. 21.1 %) is shown in violet together with UlaG (in green) due to their high similarity and because of the completeness of its structure as opposed to UlaG's. Structure overlays are sorted by overall structure similarity (TM-score, Supplementary Table 4).

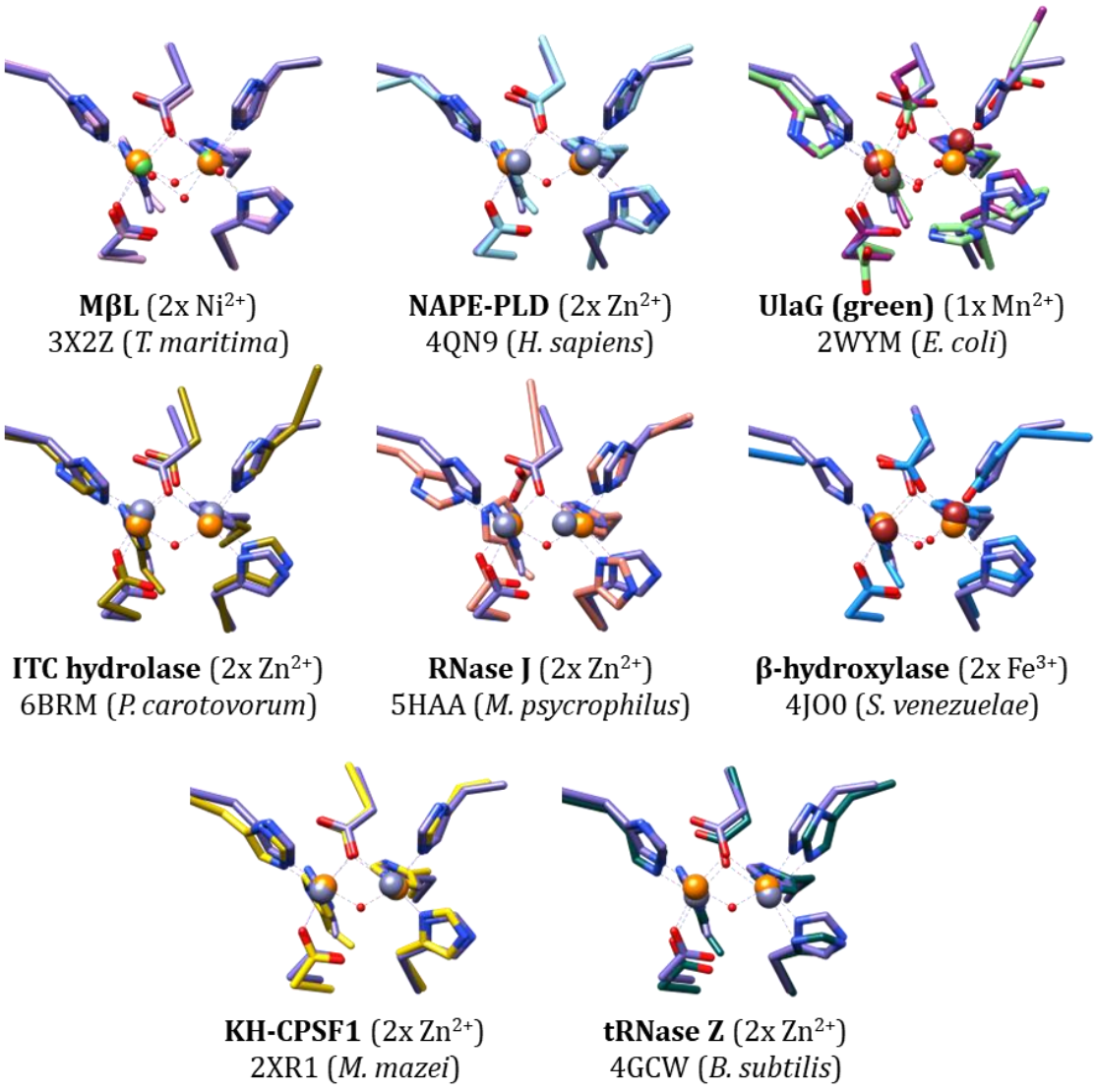


Figure 24: Metal coordination within specialized MβLs.
Igni18 (purple, two Zn ions in orange) is superimposed to all of them. Other Zn ions are depicted in light purple, Ni in green, Mn in grey and Fe in red. The uncharacterized 3BV6 protein from *V. cholerae* (2x Fe^{3+}) is shown in violet together with UlaG (in green) due to their high overall structural similarity (see Figure 23).

Protein Variable Regions (PVRs) were defined as sections or domains between two structurally aligned positions with RMSD values smaller than 2 Å (or one position in case of the N- or C-terminal domains) and a range of length variation within all structures bigger than ten amino acids. A total of ten PVRs were described for the selected structures

(Supplementary Table 5), including both N- and C-terminal ends (PVR1 and PVR10), three loops around the catalytic pocket (PVRs 2, 5 and 8), one between β9 and β10 (PVR6) and α1 (PVR3), α4 (PVR4), α5 (PVR7) and α7-α8 (PVR9; Figures 25 and 26). A histidine needed for metal coordination (His118 from Igni18, Figures 9C and 25) is located within PVR5 and is mutated in the Fe-binding enzyme structures 3BV6 and 4JO0 (Figures 23 and 24). All β-strands and α2, α3 and α6 seem to be extremely conserved.

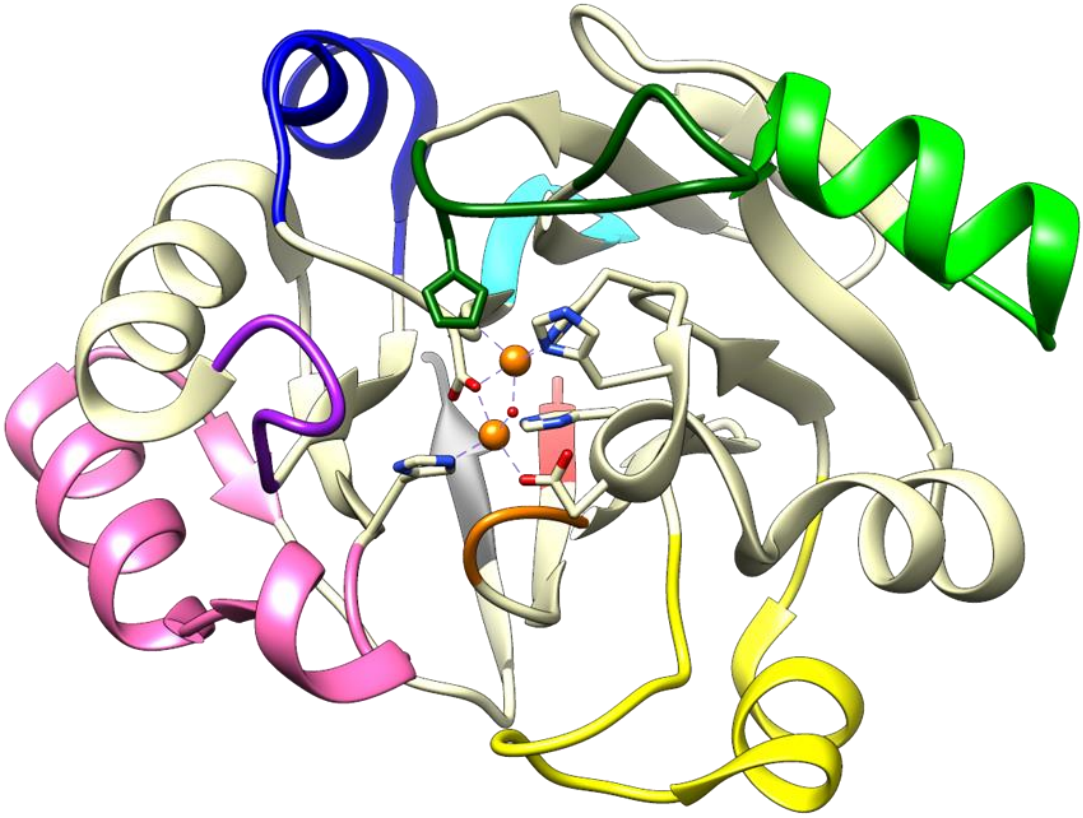


Figure 25: Protein Variable Regions (PVR) within a monomer of Igni18.
A total of 10 PVRs were defined for a subset of MβLs (Figure 23) and are shown on the structure of Igni18. The amino acids comprising each PVR and its length in number of amino acids are indicated underneath. Note that His118, involved in metal coordination is located within PVR5 (dark green).

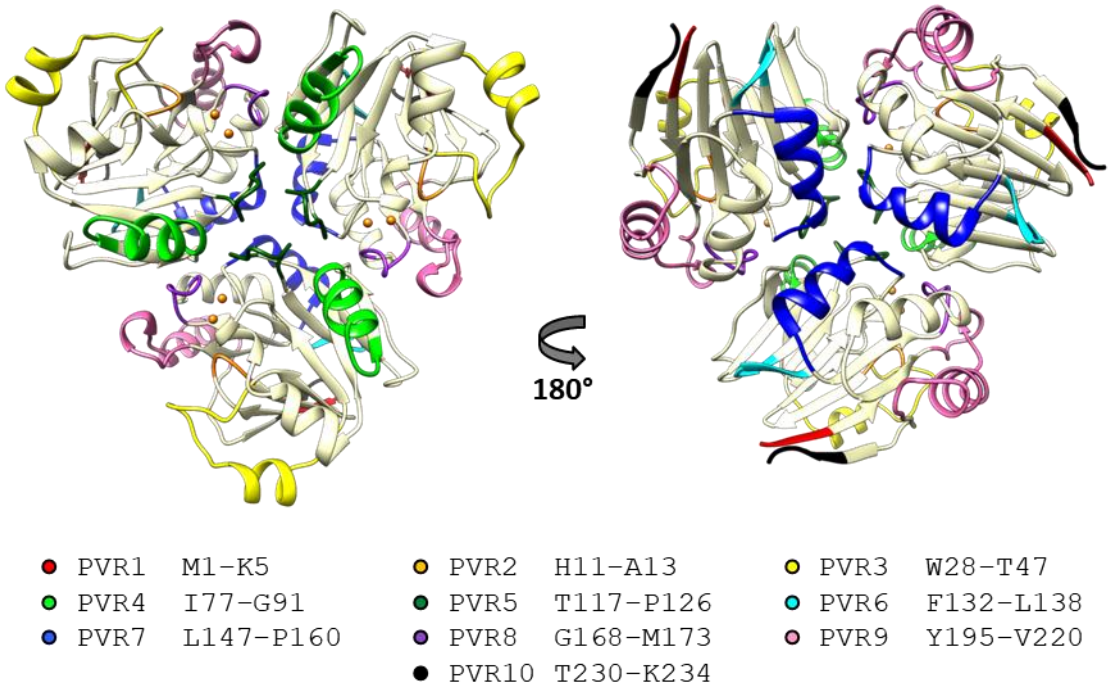
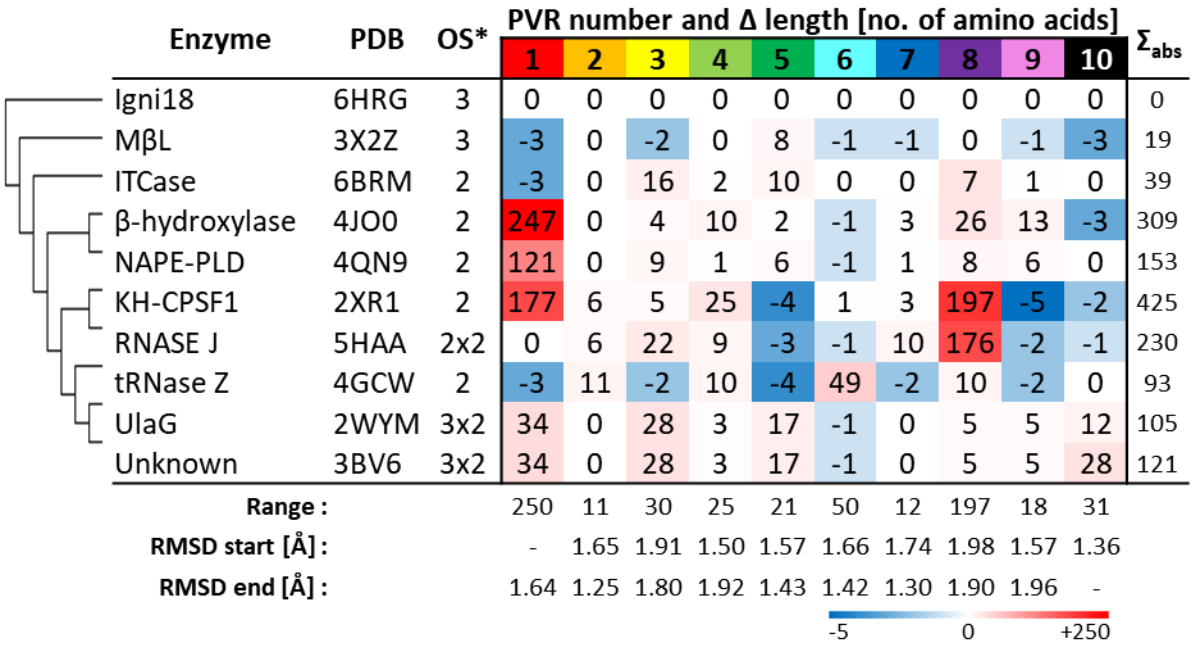


Figure 26: Protein Variable Regions (PVR) within a trimer of Igni18.
The amino acids comprising each PVR are indicated underneath. Nine out of fourteen amino acids needed for stabilization of the trimer are found within PVRs 4, 5, 7 and 8 (light-green, dark-green, blue and purple), leading to an evolutionary fast loss of this quaternary structure.

PVR lengths in number of amino acids were calculated for every structure of the subset and compared to those of Igni18 (Supplementary Table 5 and Figure 27). All PVRs revealed variability between the different enzymes, however the least diversity was found for PVR2, PVR6, PVR7 and PVR10. The MβL from *T. maritima* revealed high resemblance in PVR length for all its domains but PVR5. With eight amino acids more conforming two extra β-sheets (Figures 21 and 23), it accounts for the main difference between the two structures. The main additions of the ITCase are found in PVR3, PVR5 (+2 α-helixes) and PVR8, covering the access to the catalytic pocket, particularly by the lid-like structure formed from PVR3. The β-hydroxylase adds a large domain in front of the active site (PVR1 with +247 amino acids). It comprises at least seven β-sheets and thirteen α-helixes. Other

remarkable modifications in the PVRs limit even more the access to the active site by adding one big loop (PVR4), two α-helixes (PVR8) and one α-helix (PVR9). The structure of the NAPE-PLD adds 121 amino acids at the N-terminus (PVR1) and contains at least two α-helixes. Again, a lid-like domain comes out of PVR3, but minor than for the latter enzyme. Slight additions in PVR5, PVR8 and PVR9 also contribute to a small delimitation of the catalytic site. The archaeal CPSF1 enzyme has a 177 amino acid long N-terminal PVR1 (Figures 23 and 27). It contains two K-homology-domains [KH-domains (Mir-Montazeri *et al.* 2011)]. These are involved in RNA- or ssDNA-binding and are mainly found in proteins associated with transcriptional and translational regulation. Remarkable is as well the domain that comes out of PVR8 (+197 amino acids; Figures 23 and 27) and covers the active center completely. This β-CASP domain consists of six β-sheets and nine α-helixes and is found in pre-mRNA 3'-end-processing endonucleases (Clouet-d'Orval *et al.* 2015). PVR4 adds one α-helix and extends α4. PVR5 and PVR9 are reduced, possibly facilitating the accommodation of nucleic acids. The RNase J is similar to the CPSF1 enzyme but lacks KH-domains and the β-CASP domain contains only seven β-sheets and five α-helixes. PVR5 and PVR9 are reduced as well, while PVR3 and PVR4 are considerably larger than Igni18's. RNase Z has the only significant variation in PRV6 (Figures 23 and 27), a pre-tRNA binding domain (exosite) composed of three α-helixes (Li de la Sierra-Gallay *et al.* 2005). PVR2, which is only enlarged within the RNAses, is even larger than the ones from CPSF1 and RNase J. PVR4 and PVR8 are ten amino acids larger than the ones of Igni18. Lastly, UlaG and 3BV6 present elongations at almost every PVR, but most remarkably on PVR1, PVR3, PVR5 and PVR10 (Figure 27).



Enzyme	PDB	OS*	PVR number and Δ length [no. of amino acids]										Σ_{abs}
			1	2	3	4	5	6	7	8	9	10	
Igni18	6HRG	3	0	0	0	0	0	0	0	0	0	0	0
MβL	3X2Z	3	-3	0	-2	0	8	-1	-1	0	-1	-3	19
ITCase	6BRM	2	-3	0	16	2	10	0	0	7	1	0	39
β-hydroxylase	4JO0	2	247	0	4	10	2	-1	3	26	13	-3	309
NAPE-PLD	4QN9	2	121	0	9	1	6	-1	1	8	6	0	153
KH-CPSF1	2XR1	2	177	6	5	25	-4	1	3	197	-5	-2	425
RNASE J	5HAA	2x2	0	6	22	9	-3	-1	10	176	-2	-1	230
tRNase Z	4GCW	2	-3	11	-2	10	-4	49	-2	10	-2	0	93
UlaG	2WYM	3x2	34	0	28	3	17	-1	0	5	5	12	105
Unknown	3BV6	3x2	34	0	28	3	17	-1	0	5	5	28	121
Range :			250	11	30	25	21	50	12	197	18	31	
RMSD start [Å] :			-	1.65	1.91	1.50	1.57	1.66	1.74	1.98	1.57	1.36	
RMSD end [Å] :			1.64	1.25	1.80	1.92	1.43	1.42	1.30	1.90	1.96	-	

Figure 27: Protein Variable Region (PVR) evolution among ten MβLs.
Structural relation (left) and absolute variation in the number of amino acids (Σ_{abs}, right) compared to Igni18 as well as statistic data on PVR descriptors (total length range > 10 amino acids, start/end position RMSDs for all aligned structures < 2 Å; bottom) are also given. *Oligomerization state of the crystallized enzyme (3: trimer; 2: dimer; 2x2: double dimer; 3x2: triple dimer). Figure from Pérez-García *et al.* (submitted).

3.2 Discussion

In this work, I demonstrated that recombinant Igni18 is extremely thermostable and displays surprisingly high substrate promiscuity. Its structure resembles the archetype of an ancestral enzyme and is a very suitable model to study structural evolution within the MβL family. These three points will be discussed hereunder.

3.2.1 Igni18, hydrothermal vents and the origin of life

The Crenarchaeon *I. hospitalis* was isolated from a hydrothermal system close to Iceland (Paper *et al.* 2007). These habitats present comparable parameters to the ones found at the origin of life (Martin *et al.* 2008). *I. hospitalis*, growing anaerobically at temperatures close to the boiling point of water, could have resisted these primordial conditions. Organisms growing optimally above 80 °C are referred to as hyperthermophilic (Vieille and Zeikus 2001). For an organism to live under these harsh conditions, cellular components, including enzymes, must be very heat resistant. The MβL Igni18 showed its highest activities at temperatures between 70 °C and 95 °C (Figures 12 and 19), proving to be adapted to the living conditions of its host.

Some of the factors that have been attributed to confer protein stability at extreme temperatures are amino-acid composition, particularly a decrease in thermolabile residues, hydrophobic contacts, aromatic interactions, ion pairs and increased salt-bridge networks, oligomerization and inter-subunit interactions, packing and reduction of solvent-exposed surface area, flexibility of surface-exposed loops, and metal binding or substrate stabilization (Unsworth *et al.* 2007, Kovacic *et al.* 2016). Among others, Igni18 forms a trimer stabilized by 33 molecular interactions (3x11 hydrogen-bonds), presents a rather globular structure and binds metal ions (Figures 9 and 10). NanoDSF experiments showed

Igni18's ability to withstand high temperatures of up to 95 °C (Figure 18). Incubation at 99 °C for short times did not affect its activity (Figures 16 and 18). Incubation at 70 °C results even in an increase in activity compared to the protein stored on ice (Figure 19). Some enzymes, including α-amylases or α-glucosidases from archaeal origin (*Thermococcus* sp. and *Pyrococcus* sp.), have been shown to exhibit optimal temperatures of 100 °C to 115 °C (de Miguel Bouzas *et al.* 2006). Igni18, although active at temperatures up to 95 °C, was shown to work optimally at 90 °C. Nonetheless, it is, to the best of my knowledge, the most thermostable MβL characterized so far.

3.2.2 An extraordinary and unexpected substrate promiscuity

In this study, Igni18 was shown to convert 20 substrates (including α-naphtyl-acetate used for the zymogram in Figure 16). Thereby, the enzyme has shown activity as β-lactamase (EC 3.5.2.6), carboxylesterase/lipase (EC 3.1.1.1), PLC (EC 3.1.4.3), PDE (EC 3.1.4.1) and PTE (EC 3.1.8.1; Figures 11 and 13 and Supplementary Figures 1 and 2). However, it is uncertain if all reactions are happening at the same time inside the cell or if the processes are somehow directed within *I. hospitalis*. As the enzyme acts strongly metal-dependent, metal incorporation could be a way to steer the different activities.

The crystal structure of Igni18 revealed two metal ions bound per monomeric unit (Figure 9) and ICP-MS analysis confirmed that both are most likely Zn^{2+} (Figure 10). Nevertheless, our data imply that Igni18 can incorporate other ions as well (Figures 10 and 13). Activity assays using an array of *p*NP-substrates revealed a clear cofactor-dependency for most of its activities. The highest activities were measured in the presence of Ni^{2+} or Zn^{2+}. When supplemented with Ni^{2+}, Igni18 was most active as PDE (on bis-*p*NPP and *p*NPPP) and PLC (on *p*NP-PC; Figure 13), but also showed remarkable hydrolytic activity on

a wide range of ß-lactam antibiotics including penams and carbapenemes (Figure 11). Loaded with Zn^{2+}, Igni18 displayed β-lactamase activity on all antibiotic families tested (Figure 11) and PTE activity on paraoxon and parathion (Figure 13). In conjunction with other metals, some of the named activities can also be measured at lower rates (Figure 13). Even if metals other than Zn and Ni do not appear to be main cofactors of Igni18, it cannot be neglected that further, yet unknown activities are displayed in their presence. Lipase activity was shown not to be strictly metal-dependent (Figure 13). However, some metal ions were found to have an enhancing (Mn^{2+}, Cu^{2+}, Zn^{2+}) or a diminishing effect (Ni^{2+}) on this activity.

It is possible that *I. hospitalis* uses both versions of the enzyme for different purposes. Zn^{2+} was shown to have a much higher binding affinity than Ni^{2+} (Figure 15) and could be thought to be the main cofactor. Zn^{2+}, together with $Cu^{2+,}$ is known to have the highest binding affinity to metallo-proteins. Nonetheless, most of these metal-binding preferences do not match the metal requirements of the respective proteins (Foster *et al.* 2014). The metal composition of the deposits at the Kolbeinsey ridge was studied and interestingly, Zn^{2+} does not appear within the measured metals whereas Ni^{2+} is one of the most abundant (Lackschewitz and Wallrabeadams 1991). In the case of *I. hospitalis* using some sort of "metal-switch" to modulate the activities of Igni18, it is possible to think that the most abundant species under *in vivo* conditions would be the Ni^{2+}-loaded Igni18, which would be quickly replaced by the Zn^{2+}-containing form whenever this metal is available. It is also possible (and most probable to happen) that the organism regulates the selective import of metals during Igni18 translation with the specific metal transporters known to exist in this archaeon (Giannone *et al.* 2011, Giannone *et al.* 2015).

For some of the activities discovered, the optimal temperature and pH were determined. Interestingly, different activities showed different spectra (Figure 12). Temperature cannot be controlled by the organism. The internal pH of *I. hospitalis* is to my knowledge unknown, but it is likely to be close to neutral (Slonczewski *et al.* 2009). A slight "pH-switch" could possibly modulate the activity of Igni18, although a certain degree of pH homeostasis in the cytoplasm is assumed.

Remarkable is Igni18's degradation of phosphotriesters. These substances are inhibitors of PDEs that do not occur naturally. Therefore, it is highly interesting how Igni18 could potentially convert a phosphotriester into a phosphomonoester with the subsequent incorporation of Zn^{2+} and Ni^{2+}. Structurally different lactonases like SsoPox and SacPox from *Saccharolobus* (formerly *Sulfolobus) solfataricus* and *Sulfolobus acidocaldarius* also showed activity on phosphotriesters with Zn^{2+} as cofactor (Merone *et al.* 2008, Restaino *et al.* 2018). Phosphotriesterases could be used for the decontamination of organophosphate-polluted environments, a possible biotechnological application of Igni18.

Concerning the β-lactamase activity, it is possible that degradation of antibiotics is a side activity because of Igni18´s promiscuity. Still, there are bacteria within hydrothermal vents that produce β-lactam-antibiotics (Tasiemski *et al.* 2014, Shi *et al.* 2017) and it is possible that *I. hospitalis* defends itself against these compounds.

The activity characterization of Igni18 was performed with generic *p*NP-substrates that mimic the natural substrates with the same chemical bond that is being hydrolyzed. Considering that the *p*NP-residue can be cleaved from long-chained fatty acid esters, phosphodiesters and phospholipids, Igni18 could play a role in various pathways of *I. hospitalis* involving these chemical bonds.

3.2.3 Igni18 or how the early MβLs could have looked like

Igni18 is the deepest branching archaeal MβL enzyme characterized to date. Only 6 thermostable MβL from archaeal origin are structurally characterized (Table 11, PDB codes: 3ADR, 3AF5, 2XR1, 2YCB, 5HAA and 2VC5). A high structural similarity between the MβL from *T. maritima* (3X2Z) and Igni18 was observed, but these two organisms have more in common. *T. maritima* is a deep branching, hyperthermophilic bacterium belonging to the group of Thermotogales. Its genome of 1.86 Mbp (approx. 1,900 genes) is not much larger than *I. hospitalis*' (1.3 Mbp). Interestingly, around 11 % of *T. maritima*'s genome consists of archaeal genes (Zhaxybayeva *et al.* 2009). As in *I. hospitalis*, the genome organization is very compact, and transcription is barely regulated (Latif *et al.* 2013). The MβL 3X2Z is annotated as UlaG in *T. maritima*, but other Ula-proteins like UlaA, UlaB or UlaC are missing, also in *I. hospitalis*, and further, no evidences for the L-ascorbate biosynthesis pathway in these organisms exist. UlaG is supposed to be involved in the lactam ring cleavage of L-ascorbate 6-phosphate to 3-dehydro-L-gulonate 6-phosphate (Campos *et al.* 2004). Although there are structural similarities between Igni18 and the MβL UlaG (Figure 23), there are also major differences (Figure 27). In contrast to Igni18 and most other MβLs, UlaG is depending on one Mn^{2+} ion and is active in a hexameric form (Figures 23 and 27), although a side-activity on bis-*p*NPP has been shown (Garces *et al.* 2010). The UlaG family represents a group that has specified its catalytic function to become a modern descendant of an ancient predecessor (Fernandez *et al.* 2011). It is believed that ancient proteins evolved from higher to lower thermostability and from promiscuity to specificity (Wheeler *et al.* 2016).

In order to systematically describe Protein Variable Regions (PVRs) prone to evolution from the ancient MβL-fold to the more specialized forms, a scoring system was developed. Ten PVRs have been identified and analyzed for a subset of MβLs and describe the main structural differences leading to the modulation of activity and specificity. As an example, elongation of PVR2 and reduction of PVR5 and PVR9 seem to have occurred as an early adaptation for RNA hydrolysis, since they are common features of KH-CPSF1, RNase J and tRNase Z (Figure 27). Consequently, modular addition of specialized PVRs to the Igni18 backbone should lead to chimeric enzymes with the expected functions and structures very similar to those shown in Figure 23. According to Figure 27, trimerization is suggested to be the initial stoichiometric state for MβLs that would have then been replaced by single and eventually double (RNase J) or triple (UlaG) dimerization (Garces *et al.* 2010, Zheng *et al.* 2017). Nine out of fourteen amino acids needed for trimerization of Igni18 and the *T. maritima* MβL reside within PVR4, PVR5, PVR7 and PVR8, which might explain the early loss of the original quaternary structure (Figure 26).

Overall, the activity profile of Igni18 and the genomic analysis of *I. hospitalis* fit together in the so-called "Patchwork Hypothesis" (Fani 2012) claiming that ancestral organisms needed to function with a limited proteome and thus, enzymes with a broad specificity improved the metabolic capabilities (Ycas 1974, Jensen 1976). Primitive, meaning relatively slow and unspecific, enzymes contributed to several different metabolic pathways as a possibility to compensate for the small genome. Typically, these enzymes have low catalytic rates, like Igni18 (Figure 14). The evolution of new enzymatic functions can take place by gene duplication and subsequent functional divergence (Figure 28). The MβL-fold can be

considered as ancient predecessor of which new catalytic activities have evolved (Fernandez *et al.* 2011).

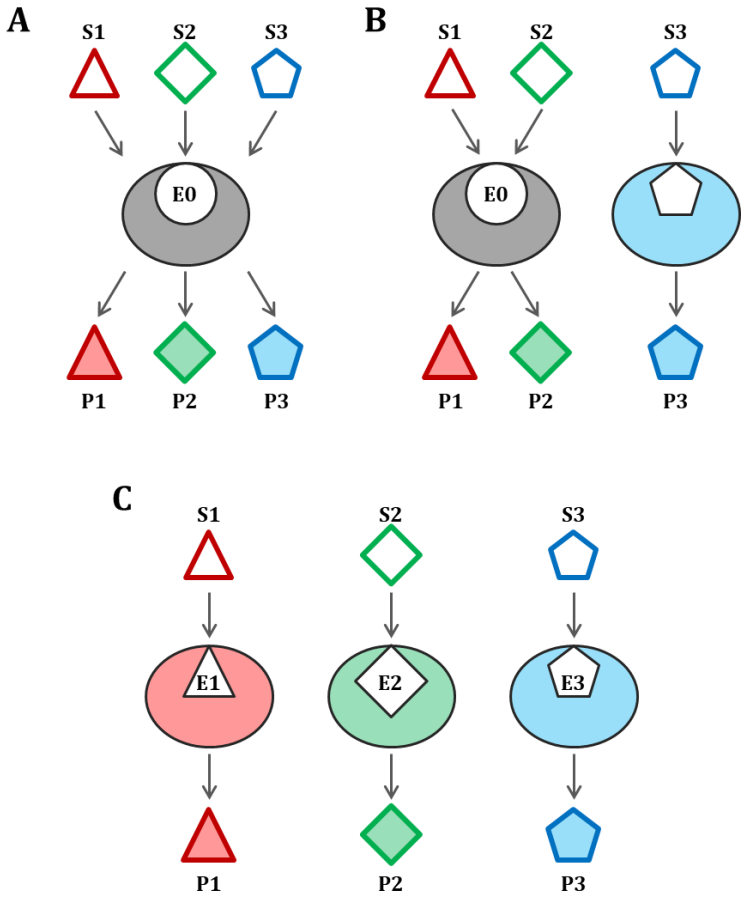


Figure 28: Patchwork hypothesis - from ancestral promiscuity to modern specialization.
A primordial enzyme (E0) is involved in the conversion of substrates S1, S2 and S3 to the products P1, P2 and P3 (A). After gene duplication, one copy evolves to specifically accept S3 (*e.g.* by adding new loops on a PVR) while E0 remains conserved (B). After a successive gene duplication, the three original enzymes can independently adapt to one of the substrates and so the modern E1, E2 and E3 are generated from the primordial E0 (C). Figure adapted from (Fani 2012).

Sequence- and structure-based analyses of the molecular evolution of UlaG show that these enzymes´ predecessors were RNA metabolizing enzymes [Figure 27 left, (Garces *et al.* 2010)]. In fact, the genomic surrounding of *igni18* suggests this, too. Directly downstream of *igni18* is a predicted gene encoding a TatD-related DNase (endodeoxyribonuclease

producing 5´-phosphomonoesters; Igni_1253; ABU82429.1), a gene encoding PurL1 for purine biosynthesis (phosphoribosylformylglycinamidine synthase II; Igni_1252; ABU82429.1) and an ATP-dependent RNA ligase (Igni_1251; ABU82427.1, Figure 29).

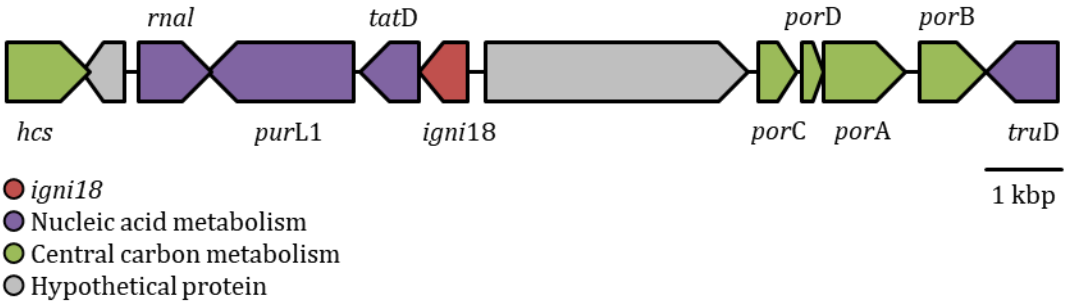


Figure 29: Genetic surrounding of *igni18*.
hcs: Homocitrate synthase, *rnal*: RNA ligase, *pur*L1: Phosphoribosylformylglycinamidine synthase, *tat*D: Trans-Activator of Transcription D (Mn-dependent DNase activity), *por*: Pyruvate oxidoreductase, *tru*D: tRNA pseudouridine synthase.

Altogether, these data imply that Igni18 is one of the best studied ancestral and deep branching hydrolases. Because of its promiscuous activities on β-lactams, it represents the first example and paradigm of archaeal MβL enzymes. Further data obtained here will give us a better understanding on the evolution of this important enzyme family. In the long run, this work could help us to design better antibiotics that are resistant to MβL hydrolysis or even inhibitors for MβLs.

IV. ENZYME PROMISCUITY WITHIN THE A/B-HYDROLASE SUPERFAMILY

IV. Enzyme promiscuity within the α/β-hydrolase superfamily

After discovering that modern enzymes most probably evolved from a promiscuous ancestor, this thesis wanted to identify the traits leading to enzyme promiscuity. The second part of this thesis starts with a collection of 142 TBT-degrading clones from 8 metagenomic libraries (referred to as the "lipase toolbox"). By means of a high-throughput screening method, the ability of every clone to degrade a large collection of esters was quantified. Single, putative esterase gene sequences were inferred from the metagenomic inserts via a sequence-based metagenomic approach. They were then cloned and tested for activity, an initial characterization step for their possible biotechnological application. Some of these results were published in (Martinez-Martinez *et al.* 2018).

4.1 Results

4.1.1 "Enzyme fingerprint" of putative esterase-carrying clones

The fosmid/cosmid-carrying clones in the "lipase toolbox" were tested for their ability to degrade a variety of compounds. This substrate collection comprised 96 (S01 to S96) differently substituted (mainly mono-, di- or tri-) esters (*e.g.* triglycerides, lactones, phenolic compounds, *etc.,* Supplementary Figure 3). Figure 30 gives an overview of the structural diversity.

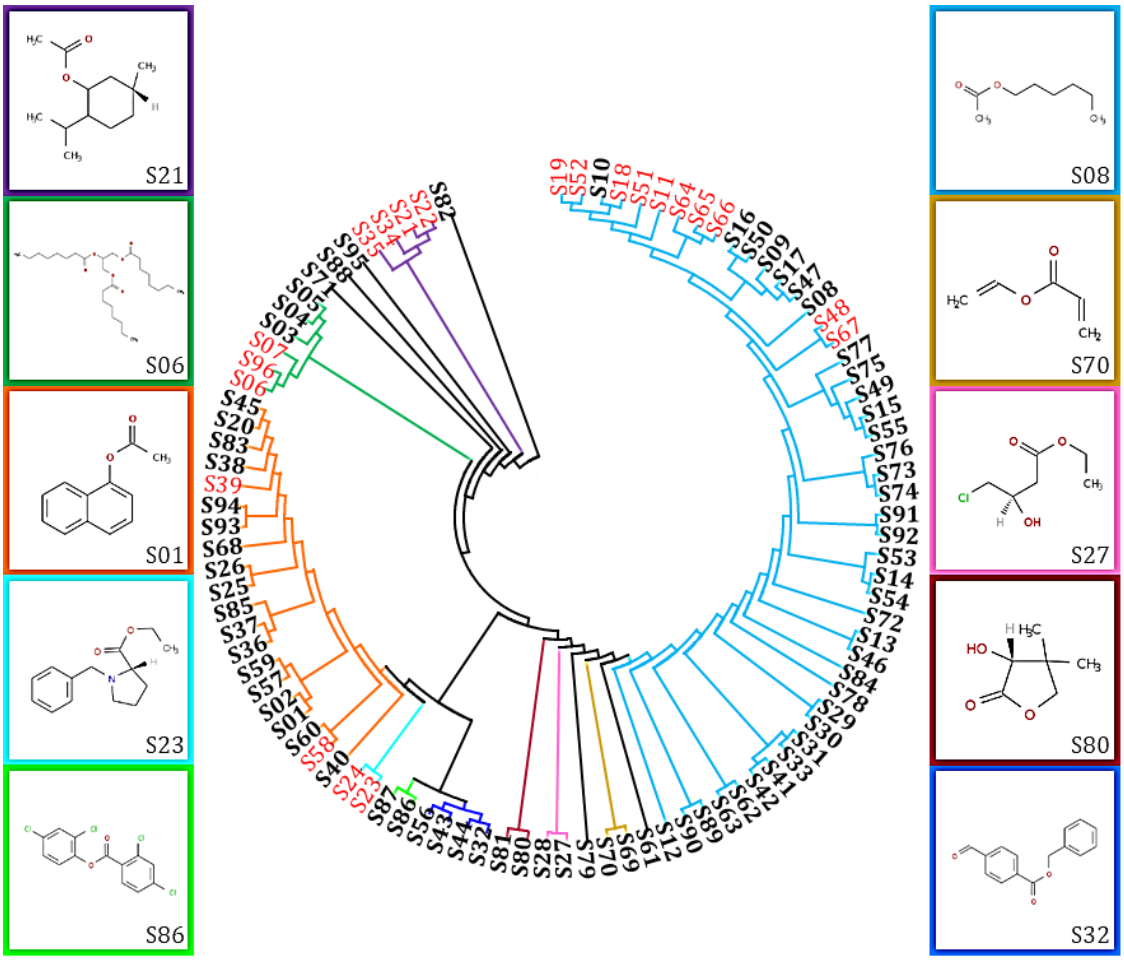


Figure 30: Hierarchical structural clustering of 96 esters.
The analysis includes all compounds used for the "fingerprint" analyses. The SMILES codes of the compounds were uploaded to the ChemMine server (Backman *et al.* 2011) and clustering was set at 0.4 similarity cutoff. The resulting Newick tree format file was visualized with MEGA X (Kumar *et al.* 2018). Each color represents a cluster. A representative structure of each group is shown. Substrates that could not be hydrolyzed by any enzyme within this study are marked in red. All structures can be found in Supplementary Figure 3.

Ester bonds are the result of the condensation of a carboxylic acid (-COOH) and an alcohol (-OH), liberating a water molecule (H_2O). The hydrolysis of an ester bond results in the restoration of both moieties. The pK_a of carboxylic acids tends to be lower than 5, this results in the deprotonation of the product at physiological pH. In a weakly buffered environment, release of H^+ will eventually result in the decrease of the pH, which can be measured spectrophotometrically in the presence of a suitable pH indicator. This principle

was applied to detect and quantify the activity of the "lipase toolbox" on all the substrates in Figure 30. This method creates a unique pattern for every esterase assayed that resembles an "enzyme fingerprint" (Figure 31).

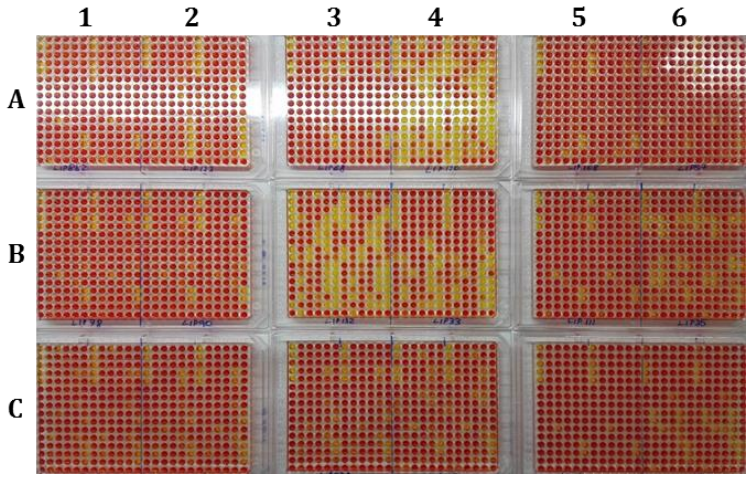


Figure 31: "Enzyme fingerprints" of 18 esterases.
Each enzyme was assayed in duplicates against 96 different esters (Supplementary Figure 3). Activity correlates with the intensity of the yellow coloration in the well. A1: Lip8b2, A2: Lip123, A3: Lip68: A4: Lip120, A5: Lip108, A6: Lip57, B1: Lip78, B2: Lip90, B3: Lip132, B4: Lip33, B5: Lip11, B6: Lip35, C1: Lip53, C2: Lip91, C3: Lip93A, C4: Lip105, C5: Lip112, C6: Lip36. Original picture by M. Ferrer ©.

The activity of all 142 clones (Lip1 to Lip142) had been confirmed in previous screening assays on TBT (S05, Figure 30 and Supplementary Figure 3) plates. Nevertheless, 81 fosmid-carrying clones did not show a measurable activity on any of the tested compounds. This phenomenon will be discussed subsequently in more depth. For the remaining 61 strains, hydrolysis of at least one substrate was observed (Figure 32). The more substrates an enzyme can convert, the more promiscuous it is (Martinez-Martinez *et al.* 2018). Remarkably, Lip120 and Lip132 showed an outstanding promiscuity as they were able to degrade 67 and 50 of the 96 tested substrates, respectively. These values account for almost 70 % of all tested compounds for Lip120.

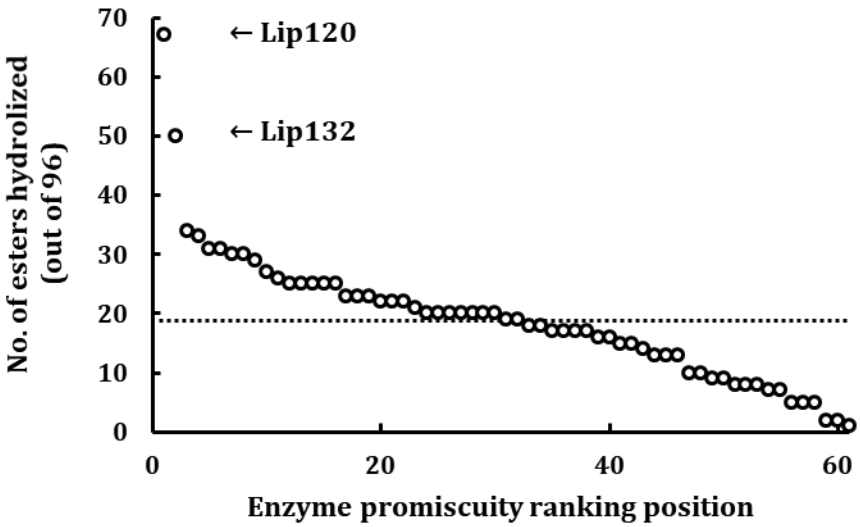


Figure 32: Ranking of esterases by promiscuity.
The 61 esterases hydrolyzing one or more esters are depicted. Dotted line indicates the median of the subset (19 hydrolyzed esters).

Figure 33 shows the normal distribution of the number of esters converted by the enzymes. Both mean and median are almost the same (19.05 and 19) indicating a good fit into a normal distribution visualized by a pointy, almost symmetrical curve with only few outliers on the right tail. In contrast, the enzymes published in (Martinez-Martinez *et al.* 2018) show a higher variability and a more dispersed distribution curve. Mean and median are further away from each other and the distribution curve is flatter and less symmetric.

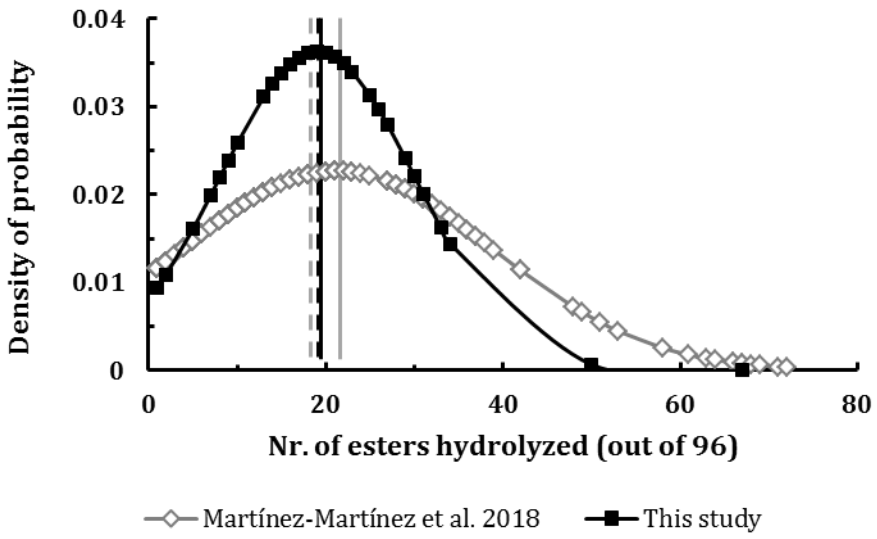


Figure 33: Normal distribution of the number of esters hydrolyzed by the esterases.
Enzymes from this study are displayed against the ones published in (Martinez-Martinez *et al.* 2018). Lines indicate mean, dotted lines median of the subsets.

A hierarchical clustering of enzymes and their ability to hydrolyze each of the substrates reveals several aspects about both the enzymes and the substrates (Figure 34). Focusing on the substrates, it appears that there are 4 different groups with increasing difficulty to being degraded. Group 1 comprises short-chained (C2, C3 and C4) substrates that seem to be very easily hydrolyzed. Most of the enzymes show activity on substrates of the Group 2, but the number decreases remarkably for Group 3. Lastly, Group 4 comprises substrates that are extremely difficult to degrade. In fact, only 18 of the 61 active clones degrade at least one of them. Enzymes can be separated in three distinct groups according to their degree of promiscuity towards specific compounds. Group A includes enzymes that can hydrolyze a minimal set of substrates and are not promiscuous. Group B is the largest group; it can hydrolyze substrates from the Groups 2 and 3, and five of them even one or two from Group 4. Enzymes in Group C can hydrolyze notably more substrates from Groups 3 and 4 and are therefore very promiscuous.

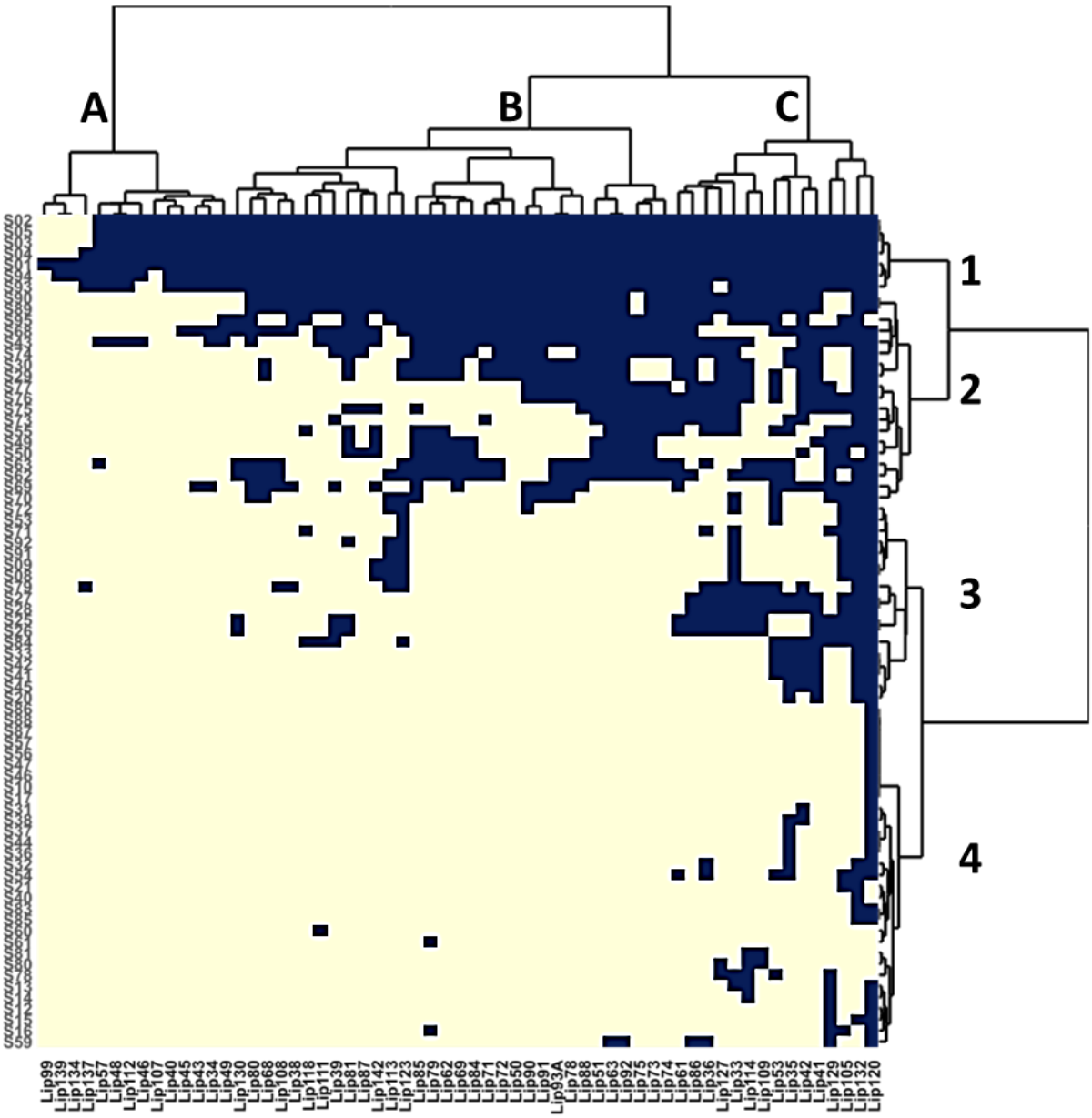


Figure 34: Hierarchical clustering of enzymes and their promiscuity.
Enzymes degrading at least one compound and substrates that were degraded by at least one of the esterases are depicted. Blue color indicates detected hydrolytic activity on the compound.

4.1.2 Assigning single protein sequences to the clones

Until this point in this study, esterase/lipase activity had only been shown for clones carrying mainly fosmids or cosmids containing metagenomic DNA. Cosmids are known to easily take up more than 10 kbp of metagenomic DNA and fosmids usually ligate around

30-40 kbp. The questions arose concerning which was the enzyme responsible for the exhibited activity, and whether a single enzyme or perhaps a combination of more was accounting for the displayed promiscuity. To assess these questions, the DNA of all 142 constructs was sequenced. In order to facilitate the handling, pools were made from the purified samples (LipH_f_1, 2, 3, and 4). After library construction, Illumina® MiSeq sequencing, quality check and assembly, four files with an outcome of almost 7 Mbp were generated. Results on assembly statistics and number of predicted ORFs are found in Table 13.

Table 13: Statistics on assembly and ORF prediction.

	LipH_f_1	LipH_f_2	LipH_f_3	LipH_f_4
Size assembled [Mbp]	1.67	1.54	1.72	1.74
Number of contigs	858	631	533	918
Mean contig size [bp]	1,946	2,441	3,227	1,895
Maximal contig size [bp]	41,639	42,860	51,756	43,599
Predicted ORFs	1,441	1,367	1,632	1,585

A CD-search (Table 10) of all ORFs and posterior filtering revealed the presence of 34 sequences with a predicted α/β-hydrolase domain, 41 putative esterases and 11 possible lipases to a total of 86 enzymes with a high probability of displaying the desired activities. The sequences were pooled and aligned with M-Coffee from the T-Coffee software package (Notredame *et al.* 2000, Wallace *et al.* 2006). Representative sequences of the eight classical lipase families (Arpigny and Jaeger 1999) as well as the later described LipS and LipT families (Chow *et al.* 2012) were included in the analysis. The resulting cladogram (Figure 35) shows molecular relations between all aligned sequences. Most of the here discovered enzymes show molecular similarities to defined lipase families. Interestingly, 25 enzymes

do not cluster within or close to any known lipase family. No discovered enzyme seems to belong to families I.1, I.2, I.4, I.5, I.6 or LipS.

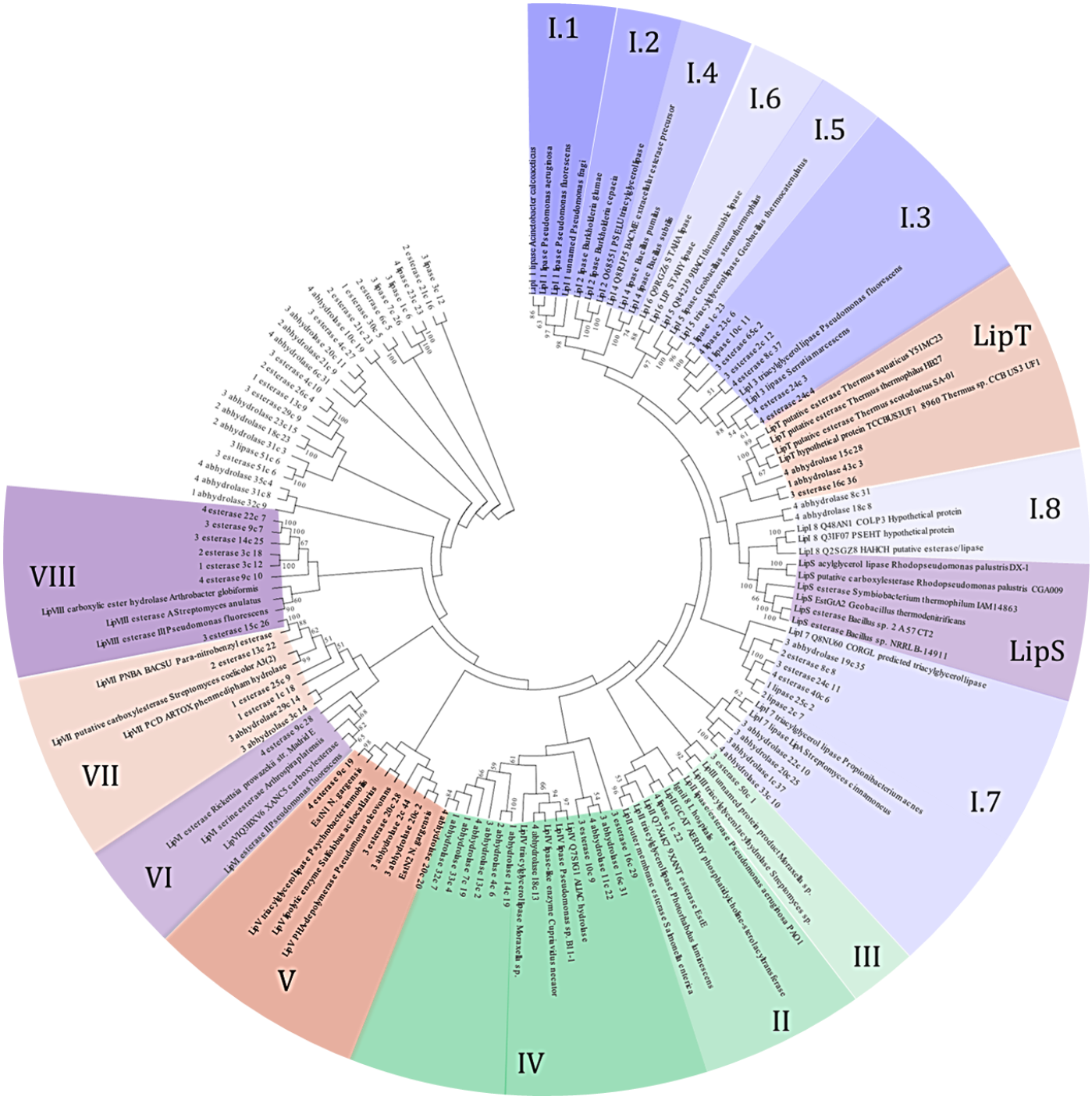


Figure 35: Molecular relations of newly-discovered hydrolases and known ABH families.
Colors highlight the representative sequences of lipase families I to VIII (Arpigny and Jaeger 1999), LipS and LipT (Chow *et al.* 2012). Closely related enzymes discovered in this study are grouped to the most similar families.

In order to assign the discovered enzyme sequences to an already known cosmid/fosmid, the following approach was followed: active clones were sequenced (Sanger method) with suitable forward and reverse primers (Table 6). After quality controlling the resulting sequences, these were compared with all contigs from the same fosmid pool (BLAST n, Table 10). In most cases, this resulted in one hit, preferably a single contig. If so, putative hydrolytic ORFs were checked if they belonged to any of the found contigs. Furthermore, both the Sanger read and the matched contig were compared against NCBI's nr-database. No BLAST hit was longer than 1.5 kbp, indicating that only single genes were matching to the deposited sequences (Supplementary Table 6). Table 14 shows the cases for which (at least) one enzyme of interest could be linked to a specific fosmid. From the 88 realized Sanger sequencings (44 forward and 44 reverse), assignment to a contig was not possible for 12 reads, meaning that 86 % of them could be correlated. Surprisingly, from the 22 contigs to which genes of hydrolytic enzymes could be associated with, almost half contained two or more hydrolase gene sequences. This was also observed for Lip120 and Lip132, which had been identified previously as the clones with the most promiscuous activities (Figures 32 and 34). Also, some fosmids/cosmids revealed the same BLAST hit within a sequencing-pool (*e.g.* Table 14: Lip086, Lip087, Lip090 and Lip091) or even matched contigs of other pools which were sequenced independently (Supplementary Table 6). This is an indication that some clones might carry the same insert and will be discussed later.

Table 14: Discovery of putative ABH-coding genes within the active constructs.
*Number N of substrates hydrolyzed out of 96. **These enzymes are found within the same ORF (LipH_f_3_51c_6) and should be counted as one.

Lip number	N/96*	LipH_f_#	Contig matched	Enz. No(s).	Enzyme name(s)
Lip041	29	1	LipH_f_1_25c	2, 6	1_lipase_25c_2, 1_esterase 25c_9
Lip051	23	2	LipH_f_2_21c	15, 16, 22	2_esterase_21c_23, 2_esterase_21c_16, 2_abhydrolase_21c_9
Lip061	25	2	LipH_f_2_6c	19	2_esterase_6c_5
Lip086	25	3	LipH_f_3_4c	42, 43	3_esterase_4c_10, 3_esterase_4c_27
Lip087	15	3	LipH_f_3_4c	42, 43	3_esterase_4c_10, 3_esterase_4c_27
Lip088	21	3	LipH_f_3_23c	39	3_abhydrolase_23c_15
Lip090	20	3	LipH_f_3_4c	42, 43	3_esterase_4c_10, 3_esterase_4c_27
Lip091	19	3	LipH_f_3_4c	42, 43	3_esterase_4c_10, 3_esterase_4c_27
Lip105	33	3	LipH_f_3_13c	33	3_abhydrolase_13c_2
Lip107	5	3	LipH_f_3_51c	29, 54**	3_lipase_51c_6, 3_esterase_51c_6
Lip108	14	3	LipH_f_3_2c	31, 41	3_abhydrolase_2c_44, 3_esterase_2c_12
Lip109	19	3	LipH_f_3_16c	34, 48	3_abhydrolase_16c_31, 3_esterase_16c_29
Lip111	13	3	LipH_f_3_15c	47	3_esterase_15c_26
Lip112	8	3	LipH_f_3_7c	28	3_lipase_7c_26
Lip113	18	3	LipH_f_3_24c	51	3_esterase_24c_11
Lip114	27	3	LipH_f_3_19c	35	3_abhydrolase_19c_35
Lip120	67	3	LipH_f_3_20c	36, 37	3_abhydrolase_20c_2, 3_abhydrolase_20c_11
Lip129	30	4	LipH_f_4_9c	60, 61	4_esterase_9c_19, 4_esterase_9c_28
Lip132	50	4	LipH_f_4_18c	75, 76	4_abhydrolase_18c_8, 4_abhydrolase_18c_13
Lip137	5	4	LipH_f_4_1c	56	4_lipase_1c_22
Lip139	2	4	LipH_f_4_15c	74	4_abhydrolase_15c_28
Lip142	18	4	LipH_f_4_20c	77	4_abhydrolase_20c_20

In silico ABH gene-construct mapping was confirmed via PCR. Single genes were amplified with specific primers and the corresponding cosmid/fosmid as template. Figure 36 shows that all tested assignments were correct.

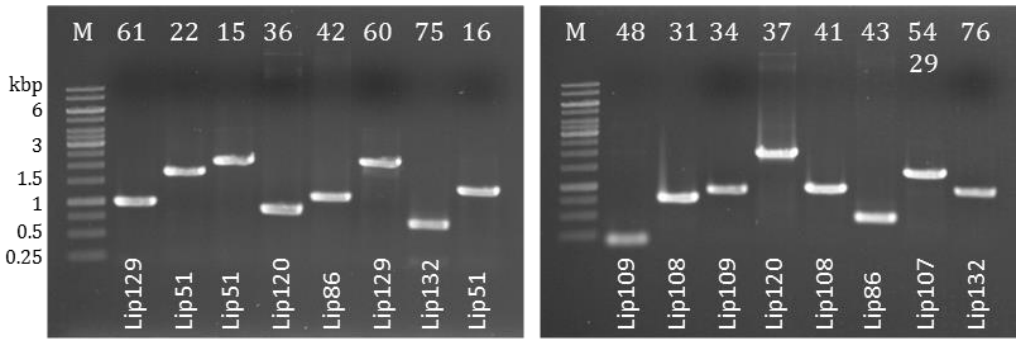


Figure 36: Specific amplification of hydrolase genes of interest.
Upper row shows the enzyme numbers amplified (see Table 14), bottom row indicates the fosmid/cosmid used as template. M: Thermo Scientific™ GeneRuler™ 1 kb DNA-Ladder. Note that 3_lipase_51c_6 and 3_esterase_51c_6 (29 and 54) are the same ORF and thus the same gene.

4.1.3 Traits leading to enzyme promiscuity within ABHs

To this point, 11 putative hydrolase-coding genes could be assigned to a unique construct. An alignment was performed and "enzyme fingerprint" data were crossed to test if enzyme specificity or promiscuity could be inferred from protein sequence. Figure 37 shows the molecular and promiscuity relations between these enzymes. Since non-promiscuous catalysts do not cluster but rather appear within different branches of the cladogram, enzyme specialization seems to have occurred at independent times in evolution. Nevertheless, some enzymes displaying low-promiscuity cluster together and might have had an already specific common ancestor (4_lipase_1c_22, 3_lipase_7c_26 and 3_esterase_15c_26; Figure 37).

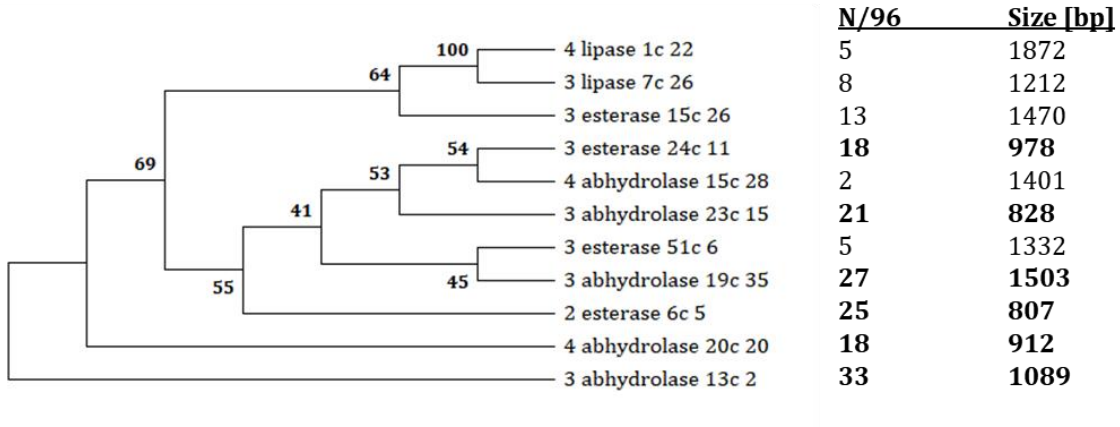


Figure 37: Relation of enzyme promiscuity to amino acid sequence and gene-length.
The most promiscuous enzyme was set at the root of the tree according to the theory of promiscuity being an ancient trait. The number of substrates degraded by each hydrolase-containing fosmid/cosmid clone during the "fingerprint"-assay (N/96; mean of the subset is 16) and gene size are indicated on the right. Enzymes degrading an above-mean number of compounds are highlighted.

Although sequence homology is not enough to correlate a protein to enzyme promiscuity, gene size seems to be able to discriminate between specific and promiscuous catalysts (Figure 38). For this analysis, the protein 3_abhydrolase_19c_35 was omitted. The enzyme shows 12 additional Met within the sequence, and it is not possible to ensure which start codon (ATG) leads to the active protein.

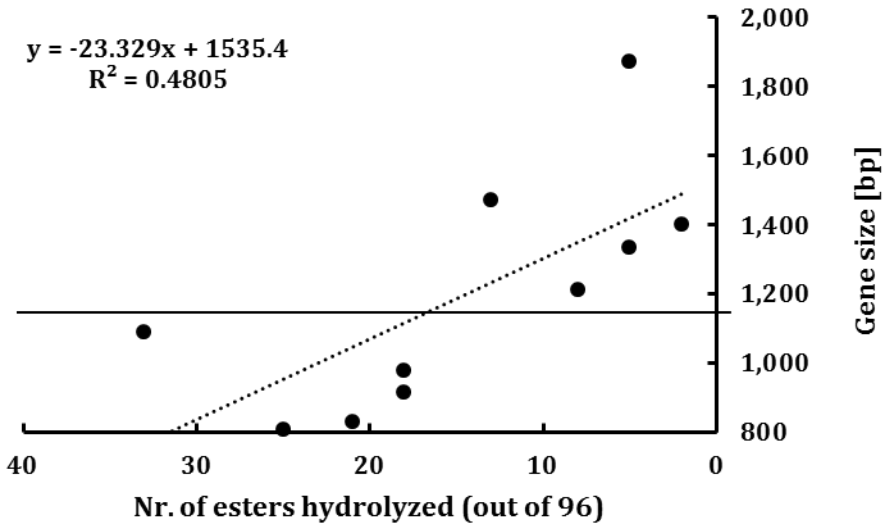


Figure 38: Possible relation of ABH gene size to promiscuity.
The analysis includes data from metagenomic clones for which a single hydrolase-coding gene could be inferred (Table 14). Dotted lines indicate linear regression; function and R^2 values are indicated at the upper-left corner. A continuous line divides the dataset in two populations: specific (>1,149 bp or 383 amino acids) and promiscuous enzymes (<1,149 bp).

According to the information gained above, the collection of predicted ABH enzymes can be divided by their gene/amino acid length into most probably specific and most probably promiscuous enzymes. From the 80 genes, 49 code for proteins longer than 383 amino acids (specific) and 31 for shorter proteins (promiscuous). Figure 39 displays the distribution of these within enzymes annotated as ABH, esterase and lipase (by their conserved domains). The ratio of putative promiscuous to specific enzymes within the lipases and esterases is close to 1:1. Interestingly, far more ABHs in the dataset are probably promiscuous (approx. 4:1).

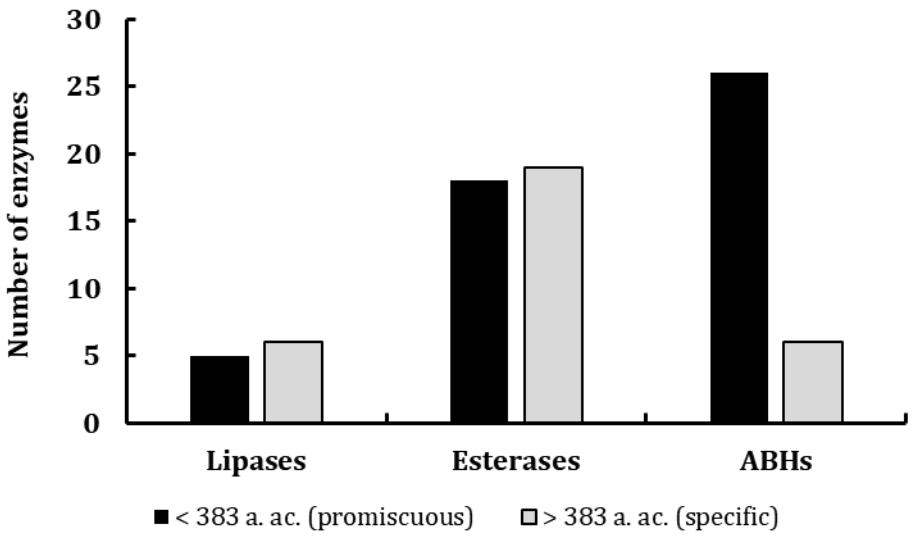


Figure 39: Putative promiscuous and specific enzymes within the "lipase toolbox"
Enzyme promiscuity was inferred from protein length in amino acids (a. ac., see Figure 38).

4.1.4 Activity verification of the single putative ABHs

The next step for the functional characterization of the 80 newly discovered enzymes was the subcloning of the respective genes into a suitable expression system. For this purpose, primers were designed to amplify and ligate the PCR products into the pDrive cloning system. Additionally, these primers contained restriction sites for subsequent directional cloning into the pET21a(+) vector and enable induction in the *E. coli* BL21 or Shuffle® T7 expression hosts under the T7 promoter (Supplementary Table 1).

72 out of the 80 genes could be cloned into the pDrive vector. Small-scale expression was performed overnight in deep-well plates, and activity was assayed against *p*NP-C8 in potassium-phosphate-Bf pH8 at 40 °C and 60 °C (Figure 40). From the 72 enzymes tested, 45 (63 %) showed esterase activity on the model substrate. The majority of the esterases displayed a higher activity at 60 °C. These can be classified as thermophilic. For esterases 8, 15, 17, 18, 49 and 75, a higher activity was measured at 40 °C. These enzymes can be

referred to as mesophilic. The three most active enzymes in this assay were 48, 56 and 79. Interestingly, 56 was linked to Lip137, which only hydrolyzed 5 out of the 96 esters, whereas 36, 37 and 75, located in the highly promiscuous Lip 120 and Lip132 exhibited rather low activities.

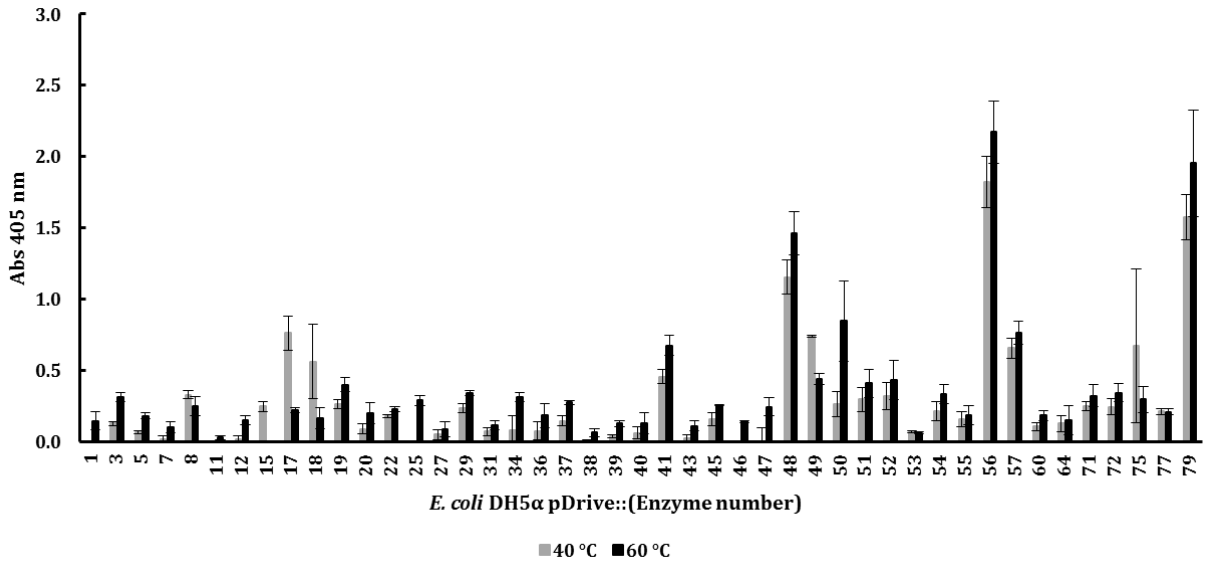


Figure 40: Activity of the subcloned hydrolases on *p*NP-C8.
The assays were performed with crude extract prepared from IPTG-induced *E. coli* cells at the indicated temperatures. Data represent mean values of triplicates and bars indicate standard error.

4.2 Discussion

During the second part of this thesis, it was shown how the application of both function- and sequence-based metagenomics can lead to the discovery of new enzymes. In this case, functional screening of seven metagenomes and enrichments lead to the identification of 142 active clones carrying putative carboxyl esterases (Table 2). Figure 41 summarizes the main procedures and findings of this work leading finally to the discovery of 80 new putative esterases, 45 of which were proven to be active on *p*NP-C8, a model esterase substrate. These and other methods and findings will be discussed hereunder.

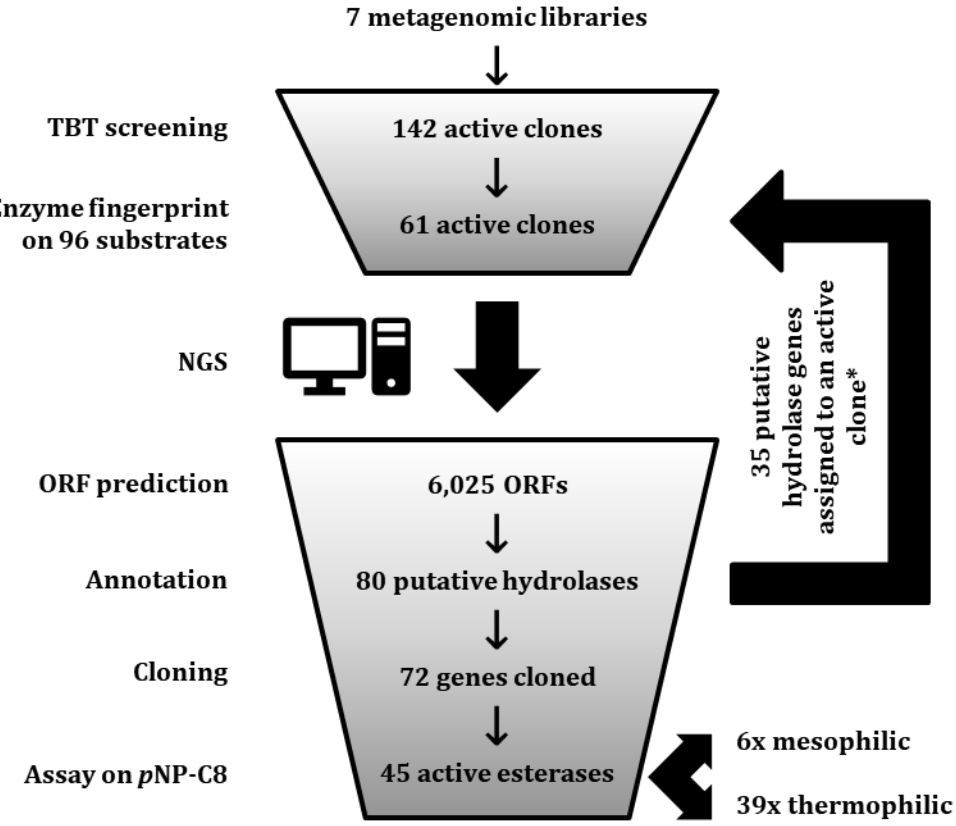


Figure 41: Screening for putative carboxyl esterases – a graphic summary.
*Some hydrolase genes were linked to more than one original metagenomic clone and each association is individually counted; 29 of them were uniquely assigned.

4.2.1 "Enzyme fingerprinting" - putting numbers to promiscuity

The most fundamental property of any catalyst is selectivity, which is defined in terms of reaction types, substrate and product range, stereochemical preferences, and operating conditions (Reymond and Wahler 2002). The first activity profile method was developed in the 1960s to characterize microorganisms for their ability to degrade an array of 16 substrates, which was further developed to the commercial product API® ZYM (bioMérieux SA, Marcy-l'Étolie, France). Since enzymes began to be applied as biotechnological tools, some of them were found to accept a range of substrates structurally different from their natural best substrate. In order to discover which types of structures and stereoselectivities are tolerated by a given enzyme, the API® ZYM concept was adapted to enzymes. Fingerprinting has been successfully adapted to many enzymes [*e.g.* proteases (Salisbury *et al.* 2002), protein kinases (Gaudet *et al.* 2003) and others (Reymond *et al.* 2009)]. For esterase prospecting, most assays monitor the pH drop caused by the liberation of one carboxylic acid during hydrolysis, which can be measured spectroscopically with a pH indicator [*e.g.* nitrophenol or phenol red; (Reymond and Wahler 2002, Alcaide *et al.* 2013)].

High-throughput screening of the 142 pre-selected clones offered a big amount of data in a reasonable time. The substrate library used comprised a large variety of different substrates, but two clusters were predominant (S01 and S08 in Figure 30). It was possible to classify the activities as specific, moderate promiscuous and prominently promiscuous by the ability to degrade a number of substrates with four levels of complexity (less than 10, 10 to 25 and more than 25, respectively; Figure 34). Most of the enzymes analyzed could degrade the short-chained esters of the lowest complexity group, but enzymes displaying high promiscuity could also accept larger substrates (Martinez-Martinez *et al.* 2018).

Remarkably, although all of them were supposed to be active at least on one of the substrates (S05, TBT), activity could only be detected for 61 clones. For the remaining 81, some of these causes may have had an influence: slow growth, low expression levels, ineffective cell lysis or the enzyme not being active under the tested conditions (*i.e.* pH, temperature, time).

4.2.2 Sequence-based screening to identify hydrolytic enzymes

Sequencing of the esterase gene-carrying constructs and downstream *in silico* screening for putative hydrolase genes revealed 80 newly described enzymes. Sequence analysis revealed a high diversity within the dataset. Most of the amino acid sequences could be assigned to eleven of the classic families (Arpigny and Jaeger 1999). However, 25 of them (approx. 31 %) remained unassigned and have the potential of being novel.

An approach involving Sanger sequencing was employed to assign nucleotide sequences to the original constructs and 35 of them could be linked to a fosmid/cosmid. Interestingly, some genes were assigned to more than one metagenomic DNA fragment. Possibly, the same clones were picked twice during library preparation or similar eDNA fragments were cloned. Furthermore, some constructs were found to carry more than one esterase-coding gene. In these cases, fingerprinting needs to be performed with clones harboring single ABH enzymes.

Initial characterization of enzymes encoded by 72 subcloned genes revealed only 45 being active. Most of them (approx. 87 %) showed a higher activity at 60 °C than at 40 °C. Thermophilic enzymes are valuable for biotechnological applications for their good kinetic stability and other advantages, *e.g.* lower medium viscosity, higher substrate solubility, higher reaction rates and fewer microbial contamination risks (Ebaid *et al.* 2019). They can

be processed under harsh conditions, which contributes to the reusability and thus reduces the overall costs (Xu *et al.* 2020).

These tests were performed with the pDrive system, which is intended to be used for cloning purposes. Nevertheless, gene expression can be induced with lactose/IPTG if the gene has been inserted in the correct direction. For a better characterization, genes should be further cloned into the pET21a(+) expression system. Interestingly, the most active enzymes were the less promiscuous. It is assumed that promiscuous, ancestral enzymes had slow conversion rates, whereas modern, specialized enzymes display higher activities (Fani 2012).

4.2.3 *In silico* discrimination of promiscuity from specificity

As a first approach to finding a pattern which can predict if an enzyme is going to react with a certain promiscuity, the number of substrates degraded during the enzyme fingerprinting were mapped to a Maximum-Likelihood tree containing the sequences of the enzymes uniquely assigned to one fosmid/cosmid. Although some specific enzymes clustered together, others could be detected grouping with promiscuous enzymes. It is known that the same trait, in this case specialization, can independently evolve more than once within an enzyme family (Alderson *et al.* 2014). Thus, molecular relations between enzymes could not predict promiscuity.

A second approach was based on the knowledge inferred from Igni18 and the MβLs. The first description of the ABH-fold pointed out the insertion of a β-hairpin loop inserted between the seventh and eighth β-strands for one of the proteins (Ollis *et al.* 1992). In the same way, as seen in Figure 27, addition of loops within PVRs has been shown to modulate the promiscuous activity of the Igni18-like MβL core towards more specific activities [*e.g.*

β-CASP domain (Clouet-d'Orval *et al.* 2015) or the exosite (Li de la Sierra-Gallay *et al.* 2005)]. In this way, one can assume that promiscuous, ancestral enzymes will lack additional modules and will have a much shorter aminoacid sequence. This simple approach proved to discriminate between specific (> 383 amino acids) and promiscuous enzymes (< 383 amino acids) for a defined subset (Figure 38), and it could be applied to the entire dataset (Figure 39). The model lipase CalB from the yeast *Candida antarctica* was shown to degrade 68 of the 96 substrates (Martinez-Martinez *et al.* 2018). With a sequence length of 321 amino acids (963 bp), it would have been identified as promiscuous by this method. Sadly, this correlation could not be shown for the dataset in (Martinez-Martinez *et al.* 2018), so another approach to identify promiscuous enzymes *in silico* had to be found – and is described below.

Data from this part of the thesis was pooled with the information gathered from enzymes from other project partners to elaborate a large study about esterase promiscuity (Martinez-Martinez *et al.* 2018). In this study, a total of 145 enzymes from diverse sources were fingerprinted. Figure 33 compares the performance of the enzymes from the Streit lab to the ones included in the publication, which covers a larger spectrum of enzyme promiscuity. Analysis of substrate range patterns in combination with structural analysis and protein–ligand simulations led to the identification of the Active Site Effective Volume (ASEV) as the major driver for protein promiscuity. ASEV is calculated as the active site cavity volume corrected by the relative solvent accessible surface area (SASA) and identifies enzyme promiscuity when above a defined threshold. This method proved to be transferrable to phosphatases, so it is likely to be applicable to other enzyme families.

V. CONCLUSIONS AND OUTLOOK

V. Conclusions and outlook

The main findings of this work on enzyme promiscuity in the MβL and ABH protein families are summarized below:

First, the MβL Igni18 of crenarchaeal origin proved to be a highly thermophilic enzyme with surprisingly high promiscuity. The enzyme revealed activity as β-lactamase (EC 3.5.2.6), esterase/lipase (EC 3.1.1.1), PLC (EC 3.1.4.3), PDE (EC 3.1.4.1) and PTE (EC 3.1.8.1). Different activities showed different temperature profiles and/or pH optima. Intriguingly, metal availability was involved in the modulation of substrate specificity. The amino acid sequence of Igni18 homologs were found to be highly conserved within archaeal species. Similar proteins could also be identified in Bacteria and Eukarya. The protein was crystallized as a trimer, each monomer binding two Zn ions. Other metals, mainly Ni, could also be bound by the protein. Igni18 could be identified as the common structure of a set of specialized MβLs. Ten Protein Variable Regions (PVRs) of the Igni18-like core domain were identified to account for the main structural differences between all analyzed proteins, which could explain the differences in multimerization patterns. Together with modifications in the metallo-chemistry of the binding pocket, PVRs are suggested to account for specialization of the enzymes towards their modern substrates. Because of its thermal stability, promiscuity, structural simplicity and wealth of biochemical data, Igni18 is proposed to exemplify the first structure of a common ancestor of these enzymes. Igni18 also represents the first example of a crenarchaeal MβL with lipase/esterase activities.

Second, a collection of 142 metagenome-derived clones displaying esterase activity was fingerprinted for its ability to degrade an array of 96 esters. The substrates converted were shown to have very diverse structures. In contrary to our initial assumption, only 61 out of the 142 clones were able to degrade one or more substrates. They could be divided into three categories by their ability to degrade several compounds as specific, moderately promiscuous and prominently promiscuous. Most of them (52 %) were assigned to the second category. In the same way, substrates from the array that could be degraded were categorized in four groups by how "easy" they were degraded by enzymes. The constructs carried by the clones were sequenced generating approx. 7 Gbp of assembled contigs. Within the approx. 6,000 predicted ORFs, 80 were annotated as ABH, esterase or lipase. Of these sequences, 55 could be assigned to known esterase families. Genes were amplified, cloned and preliminary activity tests confirmed activity of 45 enzymes. Most proteins were more active at 60 °C, making them promising candidates for biotechnological applications. Enzyme promiscuity within the ABHs could not be related to a single origin based on MSA, nor could it be inferred from protein sequence length. However, the active site effective volume was identified to be a major driver for enzyme promiscuity. This value is calculated by correcting the active site cavity volume by the relative solvent accessible surface area (SASA) and allows for *in silico* prediction. This approach was shown to also predict promiscuity in phosphatases and is thought to be transferrable to other enzyme families.

Future experiments on evolution of promiscuity of the MβLs should include more known crystal structures of MβLs to better identify PVRs. Also, more screening assays are needed to reveal how much more promiscuous the Igni18-like MβL-core can be. The results would help identify the primordial activities and distinguish them from the ones probably

acquired in the course of evolution. Likewise, the creation of chimeric enzymes could help explain which PVRs are essential to the development of new activities or the specialization of the existing ones. The deletion of *igni18* from *Ignicoccus hospitalis* would give more information on the actual function of the protein within the organism and the original function of early MβLs, but the organism is not genetically accessible to date.

As for the studies on promiscuity within the ABH-superfamily, more information needs to be collected for single enzymes related to their substrate-profiles. This could be assessed by the same fingerprint assay performed for the metagenomic clones. Protein structure predictions and analysis of the catalytic cavity would allow the calculation of the active site effective volume. Candidates displaying extraordinary promiscuity and/or other desired qualities (*e.g.* broad pH-spectrum, tolerance against heat, salinity or organic solvents) could be applied by industry. These promiscuous enzymes could be further engineered to suit these applications by shaping their cavity to increase specific activities. Furthermore, a similar structural study as the one carried out for the MβLs could help identify ancestral structures of ABHs and where and how their PVRs have been evolving.

As a final thought, these results should make us aware of promiscuity being still around us nowadays. It is important to rethink the way we have been annotating genomes and reconstructing metabolic pathways, as public databases (*e.g.* MetaCyc or KEGG) might not take enzyme promiscuity into account. Certain metabolic routes might not be optimized in some species yet, especially in deep-branching organisms with reduced genomes, and others might still have the potential to perform better. Life is not static, and evolution will always go on. These ancestral enzymes will surprise us with novel and unexpected functions to address future challenges in modern biotechnology.

VI. ACKNOWLEDGEMENTS

VI. Acknowledgements

I couldn't start this section without thanking Prof. Wolfgang R. Streit for welcoming me in his lab, for trusting my work and for giving me so much freedom to develop the topic, for the good ideas and suggestions and the motivation through the years. I thank Prof. Bernward Bisping for being my second supervisor. I'm very, very thankful to Dr. Jennifer Chow for all her support and guidance, for the long days and all the moments together, the good and the difficult ones - and for everything that wouldn't fit in these lines. I acknowledge our technical team, especially Angela and Inka, for their dedication and assistance. I am thankful for the economic support by the European Union´s Horizon 2020 research and innovation program Blue Growth (grant agreement no. 634486 INMARE).

Some great people were very important for the realization of this thesis: I would like to thank Harald Huber for providing gDNA from *I. hopsitalis* and Nico Holzscheck for starting the Igni18 madness with the cloning and the initial expression. Clemens Bernhard, for helping me upscaling the *Pichia* fermentation and at the HPLC, with Matthieu Nourrison. I want to thank Mónica Martínez-Martínez, Cristina Coscolín and Manuel Ferrer for performing the enzyme fingerprinting and writing a beautiful publication coordinating the work of so many people. Thank you, Stefanie Kobus, Astrid Höppner, Sander H. Smits and the team at the ESRF in Grenoble for the beautiful Igni18 "diamond". And Christian Strunk, Filip Kovacic and Karl-Erich Jäger for the nanoDSF experiments. A big 'thank you' to everyone who contributed to the manuscripts, for your valuable comments and enthusiasm.

To Domm, Philipp and other Erasmus people I met again, ‘gracias amigos’ for making me stay here for all this time. To the people from the lab who left and who stayed, from my group and others: thank you for making these #AKNM years amazing, for the crazy moments inside and outside of the lab/institute, for turning from colleagues into friends and so much more.

I’m also grateful to have my Hamburg family and everyone who visited me during these years, thank you for making me feel like home while being more than 2,000 km away from ‘la terreta’. To Fabian, the “most Spanish guy” from Australia, thanks for certifying the language correctness of this thesis.

My very special gratitude goes to Meri, for all her interest in my work, her support and affection.

Last but for sure not least, this is for my parents to understand:

Gracias a mi familia, por siempre darme todo lo mejor y por nunca dejar de apoyarme, aunque aún no tengáis muy claro qué hago con “las proteínas, las bacterias y las levaduras”. A mi hermana, porque no me la merezco y cree en mí más que yo mismo. Os quiero, PereZitos.

VII. REFERENCES

VII. References

Alcaide, M., P. J. Stogios, A. Lafraya, A. Tchigvintsev, R. Flick, R. Bargiela, T. N. Chernikova, O. N. Reva, T. Hai, C. C. Leggewie, N. Katzke, V. La Cono, R. Matesanz, M. Jebbar, K. E. Jaeger, M. M. Yakimov, A. F. Yakunin, P. N. Golyshin, O. V. Golyshina, A. Savchenko, M. Ferrer and M. Consortium (2015). "Pressure adaptation is linked to thermal adaptation in salt-saturated marine habitats." Environ Microbiol **17**(2): 332-345.

Alcaide, M., J. Tornes, P. J. Stogios, X. Xu, C. Gertler, R. Di Leo, R. Bargiela, A. Lafraya, M. E. Guazzaroni, N. Lopez-Cortes, T. N. Chernikova, O. V. Golyshina, T. Y. Nechitaylo, I. Plumeier, D. H. Pieper, M. M. Yakimov, A. Savchenko, P. N. Golyshin and M. Ferrer (2013). "Single residues dictate the co-evolution of dual esterases: MCP hydrolases from the alpha/beta hydrolase family." Biochem J **454**(1): 157-166.

Alderson, R. G., D. Barker and J. B. Mitchell (2014). "One origin for metallo-beta-lactamase activity, or two? An investigation assessing a diverse set of reconstructed ancestral sequences based on a sample of phylogenetic trees." J Mol Evol **79**(3-4): 117-129.

Arpigny, J. L. and K. E. Jaeger (1999). "Bacterial lipolytic enzymes: classification and properties." Biochem J **343 Pt 1**: 177-183.

Backman, T. W., Y. Cao and T. Girke (2011). "ChemMine tools: an online service for analyzing and clustering small molecules." Nucleic Acids Res **39**(Web Server issue): W486-491.

Bebrone, C. (2007). "Metallo-beta-lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily." Biochem Pharmacol **74**(12): 1686-1701.

Bernhardt, C. (2019). Untersuchungen zur Bildung von Vitamin B_{12} durch *Acetobacter pasteurianus* und zur Co-Fermentation von Tofu und Pflanzenmaterial durch *Propionibacterium freudenreichii* und *Lactobacillus reuteri* für die Anreicherung mit Vitamin B_{12}. Dr. rer. nat., Universität Hamburg.

Bertani, G. (1951). "Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*." J Bacteriol **62**(3): 293-300.

Blount, Z. D. (2015). "The unexhausted potential of *E. coli*." Elife **4**.

Böhnke, S. (2010). Untersuchungen zur Struktur und Dynamik Ammoniak oxidierender Populationen in tidebeeinflussten Lebensräumen der Elbe. Dipl., University of Hamburg.

Boratyn, G. M., C. Camacho, P. S. Cooper, G. Coulouris, A. Fong, N. Ma, T. L. Madden, W. T. Matten, S. D. McGinnis, Y. Merezhuk, Y. Raytselis, E. W. Sayers, T. Tao, J. Ye and I. Zaretskaya (2013). "BLAST: a more efficient report with usability improvements." Nucleic Acids Res **41**(Web Server issue): W29-33.

Braakman, R. and E. Smith (2014). "Metabolic evolution of a deep-branching hyperthermophilic chemoautotrophic bacterium." PLoS One **9**(2): e87950.

Callebaut, I., D. Moshous, J. P. Mornon and J. P. de Villartay (2002). "Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family." Nucleic Acids Res **30**(16): 3592-3601.

Campos, E., L. Baldoma, J. Aguilar and J. Badia (2004). "Regulation of expression of the divergent ulaG and ulaABCDEF operons involved in LaAscorbate dissimilation in *Escherichia coli*." J Bacteriol **186**(6): 1720-1728.

Carfi, A., S. Pares, E. Duee, M. Galleni, C. Duez, J. M. Frere and O. Dideberg (1995). "The 3-D structure of a zinc metallo-beta-lactamase from *Bacillus cereus* reveals a new type of protein fold." EMBO J **14**(20): 4914-4921.

Carr, P. D. and D. L. Ollis (2009). "Alpha/beta hydrolase fold: an update." Protein Pept Lett **16**(10): 1137-1148.

Choi, H., H. J. Kim, A. Matsuura, B. Mikami, H. J. Yoon and H. H. Lee (2015). "Structural and functional studies of a metallo-beta-lactamase unveil a new type of structurally encoded nickel-containing heterodinuclear site." Acta Crystallogr D Biol Crystallogr **71**(Pt 10): 2054-2065.

Chow, J. (2008). Isolierung von neuartigen thermophilen Enzymen für die Biokatalyse. Dipl., University of Hamburg.

Chow, J., F. Kovacic, Y. Dall Antonia, U. Krauss, F. Fersini, C. Schmeisser, B. Lauinger, P. Bongen, J. Pietruszka, M. Schmidt, I. Menyes, U. T. Bornscheuer, M. Eckstein, O. Thum, A. Liese, J. Mueller-Dieckmann, K. E. Jaeger and W. R. Streit (2012). "The metagenome-derived enzymes LipS and LipT increase the diversity of known lipases." PLoS One **7**(10): e47665.

Clouet-d'Orval, B., D. K. Phung, P. S. Langendijk-Genevaux and Y. Quentin (2015). "Universal RNA-degrading enzymes in Archaea: Prevalence, activities and functions of beta-CASP ribonucleases." Biochimie **118**: 278-285.

Daiyasu, H., K. Osaka, Y. Ishino and H. Toh (2001). "Expansion of the zinc metallo-hydrolase family of the beta-lactamase fold." FEBS Lett **503**(1): 1-6.

de Miguel Bouzas, T., J. Barros-Velazquez and T. G. Villa (2006). "Industrial applications of hyperthermophilic enzymes: a review." Protein Pept Lett **13**(7): 645-651.

Don, R. H., P. T. Cox, B. J. Wainwright, K. Baker and J. S. Mattick (1991). "'Touchdown' PCR to circumvent spurious priming during gene amplification." Nucleic Acids Res **19**(14): 4008.

Dong, R., S. Pan, Z. Peng, Y. Zhang and J. Yang (2018). "mTM-align: a server for fast protein structure database search and multiple protein structure alignment." Nucleic Acids Res **46**(W1): W380-W386.

Ebaid, R., H. C. Wang, C. Sha, A. Abomohra and W. L. Shao (2019). "Recent trends in hyperthermophilic enzymes production and future perspectives for biofuel industry: A critical review." Journal of Cleaner Production **238**.

El-Gebali, S., J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L. L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S. C. E. Tosatto and R. D. Finn (2019). "The Pfam protein families database in 2019." Nucleic Acids Res **47**(D1): D427-D432.

Ellens, K. W., N. Christian, C. Singh, V. P. Satagopam, P. May and C. L. Linster (2017). "Confronting the catalytic dark matter encoded by sequenced genomes." Nucleic Acids Res **45**(20): 11495-11514.

Falcone, M., G. Tiseo, A. Antonelli, C. Giordano, V. Di Pilato, P. Bertolucci, E. M. Parisio, A. Leonildi, N. Aiezza, I. Baccani, E. Tagliaferri, L. Righi, S. Forni, S. Sani, M. T. Mechi, F. Pieralli, S. Barnini, G. M.

Rossolini and F. Menichetti (2020). "Clinical Features and Outcomes of Bloodstream Infections Caused by New Delhi Metallo-beta-Lactamase-Producing Enterobacterales During a Regional Outbreak." Open Forum Infect Dis **7**(2): ofaa011.

Fani, R. (2012). "The Origin and Evolution of Metabolic Pathways: Why and How did Primordial Cells Construct Metabolic Routes?" Evolution: Education and Outreach **5**(3): 367-381.

Fernandez, F. J., F. Garces, M. Lopez-Estepa, J. Aguilar, L. Baldoma, M. Coll, J. Badia and M. C. Vega (2011). "The UlaG protein family defines novel structural and functional motifs grafted on an ancient RNase fold." BMC Evol Biol **11**: 273.

Ferrer, M., R. Bargiela, M. Martinez-Martinez, J. Mir, R. Koch, O. V. Golyshina and P. N. Golyshin (2015). "Biodiversity for biocatalysis: A review of the alpha/beta-hydrolase fold superfamily of esterases-lipases discovered in metagenomes." Biocatalysis and Biotransformation **33**(5-6): 235-249.

Foster, A. W., D. Osman and N. J. Robinson (2014). "Metal preferences and metallation." J Biol Chem **289**(41): 28095-28103.

Garces, F., F. J. Fernandez, C. Montella, E. Penya-Soler, R. Prohens, J. Aguilar, L. Baldoma, M. Coll, J. Badia and M. C. Vega (2010). "Molecular architecture of the Mn2+-dependent lactonase UlaG reveals an RNase-like metallo-beta-lactamase fold and a novel quaternary structure." J Mol Biol **398**(5): 715-729.

Gaudet, E. A., K. S. Huang, Y. Zhang, W. Huang, D. Mark and J. R. Sportsman (2003). "A homogeneous fluorescence polarization assay adaptable for a range of protein serine/threonine and tyrosine kinases." J Biomol Screen **8**(2): 164-175.

Giannone, R. J., H. Huber, T. Karpinets, T. Heimerl, U. Kuper, R. Rachel, M. Keller, R. L. Hettich and M. Podar (2011). "Proteomic characterization of cellular and molecular processes that enable the *Nanoarchaeum equitans-Ignicoccus hospitalis* relationship." PLoS One **6**(8): e22942.

Giannone, R. J., L. L. Wurch, T. Heimerl, S. Martin, Z. Yang, H. Huber, R. Rachel, R. L. Hettich and M. Podar (2015). "Life on the edge: functional genomic response of *Ignicoccus hospitalis* to the presence of *Nanoarchaeum equitans*." ISME J **9**(1): 101-114.

Giovannoni, S. J., J. Cameron Thrash and B. Temperton (2014). "Implications of streamlining theory for microbial ecology." ISME J **8**(8): 1553-1565.

Green, M. and J. Sambrook (2012). "Molecular cloning: a laboratory manual 4 edition Cold Spring Harbor Laboratory Press." New York.

Hall, T., I. Biosciences and C. Carlsbad (2011). "BioEdit: an important software for molecular biology." GERF Bull Biosci **2**(1): 60-61.

Hilbig, M. and M. Rarey (2015). "MONA 2: A Light Cheminformatics Platform for Interactive Compound Library Processing." J Chem Inf Model **55**(10): 2071-2078.

Holm, L. (2019). "DALI and the persistence of protein shape." Protein Sci.

Holmquist, M. (2000). "Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms." Curr Protein Pept Sci **1**(2): 209-235.

Holzscheck, N. (2016). Etablierung eines Expressionssystems für thermostabile archaeelle Hydrolasen. M. Sc., University of Hamburg.

Huber, H., M. Gallenberger, U. Jahn, E. Eylert, I. A. Berg, D. Kockelkorn, W. Eisenreich and G. Fuchs (2008). "A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum *Ignicoccus hospitalis*." Proc Natl Acad Sci U S A **105**(22): 7851-7856.

Hyatt, D., G. L. Chen, P. F. Locascio, M. L. Land, F. W. Larimer and L. J. Hauser (2010). "Prodigal: prokaryotic gene recognition and translation initiation site identification." BMC Bioinformatics **11**: 119.

Ilmberger, N., S. Gullert, J. Dannenberg, U. Rabausch, J. Torres, B. Wemheuer, M. Alawi, A. Poehlein, J. Chow, D. Turaev, T. Rattei, C. Schmeisser, J. Salomon, P. B. Olsen, R. Daniel, A. Grundhoff, M. S. Borchert and W. R. Streit (2014). "A comparative metagenome survey of the fecal microbiota of a breast- and a plant-fed Asian elephant reveals an unexpectedly high diversity of glycoside hydrolase family enzymes." PLoS One **9**(9): e106707.

Jahn, U., M. Gallenberger, W. Paper, B. Junglas, W. Eisenreich, K. O. Stetter, R. Rachel and H. Huber (2008). "*Nanoarchaeum equitans* and *Ignicoccus hospitalis*: new insights into a unique, intimate association of two archaea." J Bacteriol **190**(5): 1743-1750.

Jensen, R. A. (1976). "Enzyme Recruitment in Evolution of New Function." Annual Review of Microbiology **30**: 409-425.

Jeske, L., S. Placzek, I. Schomburg, A. Chang and D. Schomburg (2019). "BRENDA in 2019: a European ELIXIR core data resource." Nucleic Acids Res **47**(D1): D542-D549.

Keshri, V., A. Panda, A. Levasseur, J. M. Rolain, P. Pontarotti and D. Raoult (2018). "Phylogenomic Analysis of beta-Lactamase in Archaea and Bacteria Enables the Identification of Putative New Members." Genome Biol Evol **10**(4): 1106-1114.

Kobus, S., P. Perez-Garcia, A. Hoeppner, N. Holzscheck, F. Kovacic, W. R. Streit, K. E. Jaeger, J. Chow and S. H. J. Smits (2019). "Igni18, a novel metallo-hydrolase from the hyperthermophilic archaeon *Ignicoccus hospitalis* KIN4/I: cloning, expression, purification and X-ray analysis." Acta Crystallogr F Struct Biol Commun **75**(Pt 4): 307-311.

Köhler, H. U. (2007). Lipolytic enzymes from thermophilic enrichment cultures and metagenomes. Dipl., University of Hamburg.

Kovacic, F., A. Mandrysch, C. Poojari, B. Strodel and K. E. Jaeger (2016). "Structural features determining thermal adaptation of esterases." Protein Eng Des Sel **29**(2): 65-76.

Krohn, I. (2010). Molekulare und ökophysiologische Charakterisierung bakterieller Populationen in Fließgewässern. Dipl., University of Hamburg.

Kumar, S., G. Stecher, M. Li, C. Knyaz and K. Tamura (2018). "MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms." Mol Biol Evol **35**(6): 1547-1549.

Lackschewitz, K. S. and H. J. Wallrabeadams (1991). "Composition and Origin of Sediments on the Midoceanic Kolbeinsey Ridge, North of Iceland." Marine Geology **101**(1-4): 71-82.

Laemmli, U. K. (1970). "Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4." Nature **227**: 680.

Latif, H., J. A. Lerman, V. A. Portnoy, Y. Tarasova, H. Nagarajan, A. C. Schrimpe-Rutledge, R. D. Smith, J. N. Adkins, D. H. Lee, Y. Qiu and K. Zengler (2013). "The genome organization of *Thermotoga maritima* reflects its lifestyle." PLoS Genet **9**(4): e1003485.

Lee, J. Y., H. Chen, A. Liu, B. M. Alba and A. C. Lim (2017). "Auto-induction of *Pichia pastoris* AOX1 promoter for membrane protein expression." Protein Expr Purif **137**: 7-12.

Lefort, V., J. E. Longueville and O. Gascuel (2017). "SMS: Smart Model Selection in PhyML." Mol Biol Evol **34**(9): 2422-2424.

Leigh, J. A., S. V. Albers, H. Atomi and T. Allers (2011). "Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales." FEMS Microbiol Rev **35**(4): 577-608.

Lenfant, N., T. Hotelier, Y. Bourne, P. Marchot and A. Chatonnet (2013). "Proteins with an alpha/beta hydrolase fold: Relationships between subfamilies in an ever-growing superfamily." Chem Biol Interact **203**(1): 266-268.

Lenfant, N., T. Hotelier, E. Velluet, Y. Bourne, P. Marchot and A. Chatonnet (2013). "ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins: tools to explore diversity of functions." Nucleic Acids Res **41**(Database issue): D423-429.

Leonelli, S. and R. A. Ankeny (2013). "What makes a model organism?" Endeavour **37**(4): 209-212.

Li de la Sierra-Gallay, I., O. Pellegrini and C. Condon (2005). "Structural basis for substrate binding, cleavage and allostery in the tRNA maturase RNase Z." Nature **433**(7026): 657-661.

Lockwood, S., K. A. Brayton, J. A. Daily and S. L. Broschat (2019). "Whole Proteome Clustering of 2,307 Proteobacterial Genomes Reveals Conserved Proteins and Significant Annotation Issues." Front Microbiol **10**: 383.

Magotti, P., I. Bauer, M. Igarashi, M. Babagoli, R. Marotta, D. Piomelli and G. Garau (2015). "Structure of human N-acylphosphatidylethanolamine-hydrolyzing phospholipase D: regulation of fatty acid ethanolamide biosynthesis by bile acids." Structure **23**(3): 598-604.

Makarova, K. S., Y. I. Wolf and E. V. Koonin (2019). "Towards functional characterization of archaeal genomic dark matter." Biochem Soc Trans **47**(1): 389-398.

Makris, T. M., C. J. Knoot, C. M. Wilmot and J. D. Lipscomb (2013). "Structure of a dinuclear iron cluster-containing beta-hydroxylase active in antibiotic biosynthesis." Biochemistry **52**(38): 6662-6671.

Marchler-Bauer, A., Y. Bo, L. Han, J. He, C. J. Lanczycki, S. Lu, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng, L. Y. Geer and S. H. Bryant (2017). "CDD/SPARCLE: functional classification of proteins via subfamily domain architectures." Nucleic Acids Res **45**(D1): D200-D203.

Marchot, P. and A. Chatonnet (2012). "Hydrolase versus other functions of members of the alpha/beta-hydrolase fold superfamily of proteins." Protein Pept Lett **19**(2): 130-131.

Martin, W., J. Baross, D. Kelley and M. J. Russell (2008). "Hydrothermal vents and the origin of life." Nature Reviews Microbiology **6**(11): 805-814.

Martinez-Martinez, M., M. Alcaide, A. Tchigvintsev, O. Reva, J. Polaina, R. Bargiela, M. E. Guazzaroni, A. Chicote, A. Canet, F. Valero, E. Rico Eguizabal, C. Guerrero Mdel, A. F. Yakunin and M. Ferrer (2013). "Biochemical diversity of carboxyl esterases and lipases from Lake Arreo (Spain): a metagenomic approach." Appl Environ Microbiol **79**(12): 3553-3562.

Martínez-Martínez, M., R. Bargiela, C. Coscolín, J. Navarro-Fernández, P. N. Golyshin and M. Ferrer (2016). "Functionalization and modification of hydrocarbon-like molecules guided by metagenomics: Enzymes most requested at the industrial scale for chemical synthesis as study cases." Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids: Production of Fuels and Chemicals: 1-26.

Martinez-Martinez, M., C. Coscolin, G. Santiago, J. Chow, P. J. Stogios, R. Bargiela, C. Gertler, J. Navarro-Fernandez, A. Bollinger, S. Thies, C. Mendez-Garcia, A. Popovic, G. Brown, T. N. Chernikova, A. Garcia-Moyano, G. E. K. Bjerga, P. Perez-Garcia, T. Hai, M. V. Del Pozo, R. Stokke, I. H. Steen, H. Cui, X. Xu, B. P. Nocek, M. Alcaide, M. Distaso, V. Mesa, A. I. Pelaez, J. Sanchez, P. C. F. Buchholz, J. Pleiss, A. Fernandez-Guerra, F. O. Glockner, O. V. Golyshina, M. M. Yakimov, A. Savchenko, K. E. Jaeger, A. F. Yakunin, W. R. Streit, P. N. Golyshin, V. Guallar, M. Ferrer and C. The Inmare (2018). "Determinants and Prediction of Esterase Substrate Promiscuity Patterns." ACS Chem Biol **13**(1): 225-234.

Massant, J., J. Wouters and N. Glansdorff (2003). "Refined structure of *Pyrococcus furiosus* ornithine carbamoyltransferase at 1.87 A." Acta Crystallogr D Biol Crystallogr **59**(Pt 12): 2140-2149.

Merone, L., L. Mandrich, M. Rossi and G. Manco (2008). "Enzymes with Phosphotriesterase and Lactonase Activities in Archaea." Current Chemical Biology **2**(3): 237-248.

Mir-Montazeri, B., M. Ammelburg, D. Forouzan, A. N. Lupas and M. D. Hartmann (2011). "Crystal structure of a dimeric archaeal cleavage and polyadenylation specificity factor." J Struct Biol **173**(1): 191-195.

Mistry, J., R. D. Finn, S. R. Eddy, A. Bateman and M. Punta (2013). "Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions." Nucleic Acids Research **41**(12).

Mukherjee, S., D. Stamatis, J. Bertsch, G. Ovchinnikova, H. Y. Katta, A. Mojica, I. A. Chen, N. C. Kyrpides and T. Reddy (2019). "Genomes OnLine database (GOLD) v.7: updates and new features." Nucleic Acids Res **47**(D1): D649-D659.

Nobre, T., M. D. Campos, E. Lucic-Mercy and B. Arnholdt-Schmitt (2016). "Misannotation Awareness: A Tale of Two Gene-Groups." Front Plant Sci **7**: 868.

Notredame, C., D. G. Higgins and J. Heringa (2000). "T-Coffee: A novel method for fast and accurate multiple sequence alignment." J Mol Biol **302**(1): 205-217.

Ollis, D. L., E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, J. Schrag and et al. (1992). "The alpha/beta hydrolase fold." Protein Eng **5**(3): 197-211.

Paper, W., U. Jahn, M. J. Hohn, M. Kronner, D. J. Nather, T. Burghardt, R. Rachel, K. O. Stetter and H. Huber (2007). "*Ignicoccus hospitalis* sp. nov., the host of '*Nanoarchaeum equitans*'." Int J Syst Evol Microbiol **57**(Pt 4): 803-808.

Pauwels, K., M. M. Sanchez del Pino, G. Feller and P. Van Gelder (2012). "Decoding the folding of *Burkholderia glumae* lipase: folding intermediates en route to kinetic stability." PLoS One **7**(5): e36999.

Pellegrini, O., I. Li de la Sierra-Gallay, J. Piton, L. Gilet and C. Condon (2012). "Activation of tRNA maturation by downstream uracil residues in *B. subtilis*." Structure **20**(10): 1769-1777.

Pérez-García, P. (2016). Cloning and heterologous expression of the crenarchaeal *Ignicoccus hospitalis* KIN4/I RNA-polymerase in the methylotrophic yeast *Pichia pastoris*. M. Sc., University of Hamburg.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." J Comput Chem **25**(13): 1605-1612.

Pimentel, Z. T. and Y. Zhang (2018). "Evolution of the Natural Transformation Protein, ComEC, in Bacteria." Front Microbiol **9**: 2980.

Podar, M., I. Anderson, K. S. Makarova, J. G. Elkins, N. Ivanova, M. A. Wall, A. Lykidis, K. Mavromatis, H. Sun, M. E. Hudson, W. Chen, C. Deciu, D. Hutchison, J. R. Eads, A. Anderson, F. Fernandes, E. Szeto, A. Lapidus, N. C. Kyrpides, M. H. Saier, Jr., P. M. Richardson, R. Rachel, H. Huber, J. A. Eisen, E. V. Koonin, M. Keller and K. O. Stetter (2008). "A genomic analysis of the archaeal system *Ignicoccus hospitalis-Nanoarchaeum equitans*." Genome Biol **9**(11): R158.

Price, D. R. and A. C. Wilson (2014). "A substrate ambiguous enzyme facilitates genome reduction in an intracellular symbiont." BMC Biol **12**: 110.

Restaino, O. F., M. G. Borzacchiello, I. Scognamiglio, L. Fedele, A. Alfano, E. Porzio, G. Manco, M. De Rosa and C. Schiraldi (2018). "High yield production and purification of two recombinant thermostable phosphotriesterase-like lactonases from *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus* useful as bioremediation tools and bioscavengers." BMC Biotechnol **18**(1): 18.

Reymond, J. L., V. S. Fluxa and N. Maillard (2009). "Enzyme assays." Chem Commun (Camb)(1): 34-46.

Reymond, J. L. and D. Wahler (2002). "Substrate arrays as enzyme fingerprinting tools." Chembiochem **3**(8): 701-708.

Romanos, M. A., C. A. Scorer and J. J. Clare (1992). "Foreign gene expression in yeast: a review." Yeast **8**(6): 423-488.

Salisbury, C. M., D. J. Maly and J. A. Ellman (2002). "Peptide microarrays for the determination of protease substrate specificity." J Am Chem Soc **124**(50): 14868-14870.

Sanchez-Ruiz, J. (2017). "Resurrected ancestral proteins as scaffolds for protein engineering." Abstracts of Papers of the American Chemical Society **254**.

Schmeisser, C., H. Steele and W. R. Streit (2007). "Metagenomics, biotechnology with non-culturable microbes." Applied Microbiology and Biotechnology **75**(5): 955-962.

Schmeisser, C., C. Stockigt, C. Raasch, J. Wingender, K. N. Timmis, D. F. Wenderoth, H. C. Flemming, H. Liesegang, R. A. Schmitz, K. E. Jaeger and W. R. Streit (2003). "Metagenome survey of biofilms in drinking-water networks." Appl Environ Microbiol **69**(12): 7298-7309.

Schmid, A., J. S. Dordick, B. Hauer, A. Kiener, M. Wubbolts and B. Witholt (2001). "Industrial biocatalysis today and tomorrow." Nature **409**(6817): 258-268.

Shi, Y., C. Pan, B. N. Auckloo, X. Chen, C. A. Chen, K. Wang, X. Wu, Y. Ye and B. Wu (2017). "Stress-driven discovery of a cryptic antibiotic produced by *Streptomyces* sp. WU20 from Kueishantao hydrothermal vent with an integrated metabolomics strategy." Appl Microbiol Biotechnol **101**(4): 1395-1408.

Slonczewski, J. L., M. Fujisawa, M. Dopson and T. A. Krulwich (2009). "Cytoplasmic pH measurement and homeostasis in bacteria and archaea." Adv Microb Physiol **55**: 1-79, 317.

Streit, W. R. and R. A. Schmitz (2004). "Metagenomics - the key to the uncultured microbes." Current Opinion in Microbiology **7**(5): 492-498.

Tasiemski, A., S. Jung, C. Boidin-Wichlacz, D. Jollivet, V. Cuvillier-Hot, F. Pradillon, C. Vetriani, O. Hecht, F. D. Sonnichsen, C. Gelhaus, C. W. Hung, A. Tholey, M. Leippe, J. Grotzinger and F. Gaill (2014). "Characterization and function of the first antibiotic isolated from a vent organism: the extremophile metazoan *Alvinella pompejana*." PLoS One **9**(4): e95737.

Unsworth, L. D., J. van der Oost and S. Koutsopoulos (2007). "Hyperthermophilic enzymes - stability, activity and implementation strategies for high temperature applications." FEBS J **274**(16): 4044-4056.

van den Bosch, T. J. M., K. Tan, A. Joachimiak and C. U. Welte (2018). "Functional Profiling and Crystal Structures of Isothiocyanate Hydrolases Found in Gut-Associated and Plant-Pathogenic Bacteria." Appl Environ Microbiol **84**(14).

Vieille, C. and G. J. Zeikus (2001). "Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability." Microbiol Mol Biol Rev **65**(1): 1-43.

Villeret, V., B. Clantin, C. Tricot, C. Legrain, M. Roovers, V. Stalon, N. Glansdorff and J. Van Beeumen (1998). "The crystal structure of *Pyrococcus furiosus* ornithine carbamoyltransferase reveals a key role for oligomerization in enzyme stability at extremely high temperatures." Proc Natl Acad Sci U S A **95**(6): 2801-2806.

Wallace, I. M., O. O'Sullivan, D. G. Higgins and C. Notredame (2006). "M-Coffee: combining multiple sequence alignment methods with T-Coffee." Nucleic Acids Res **34**(6): 1692-1699.

Waters, E., M. J. Hohn, I. Ahel, D. E. Graham, M. D. Adams, M. Barnstead, K. Y. Beeson, L. Bibbs, R. Bolanos, M. Keller, K. Kretz, X. Lin, E. Mathur, J. Ni, M. Podar, T. Richardson, G. G. Sutton, M. Simon, D. Soll, K. O. Stetter, J. M. Short and M. Noordewier (2003). "The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism." Proc Natl Acad Sci U S A **100**(22): 12984-12988.

Wheeler, L. C., S. A. Lim, S. Marqusee and M. J. Harms (2016). "The thermostability and specificity of ancient proteins." Curr Opin Struct Biol **38**: 37-43.

Wrede, C., A. Dreier, S. Kokoschka and M. Hoppert (2012). "Archaea in symbioses." Archaea **2012**: 596846.

Xu, Z., Y. K. Cen, S. P. Zou, Y. P. Xue and Y. G. Zheng (2020). "Recent advances in the improvement of enzyme thermostability by structure modification." Crit Rev Biotechnol **40**(1): 83-98.

Ycas, M. (1974). "On Earlier States of Biochemical System." Journal of Theoretical Biology **44**(1): 145-160.

Zanna, L. and J. F. Haeuw (2007). "Separation and quantitative analysis of alkyl sulfobetaine-type detergents by high-performance liquid chromatography and light-scattering detection." J Chromatogr B Analyt Technol Biomed Life Sci **846**(1-2): 368-373.

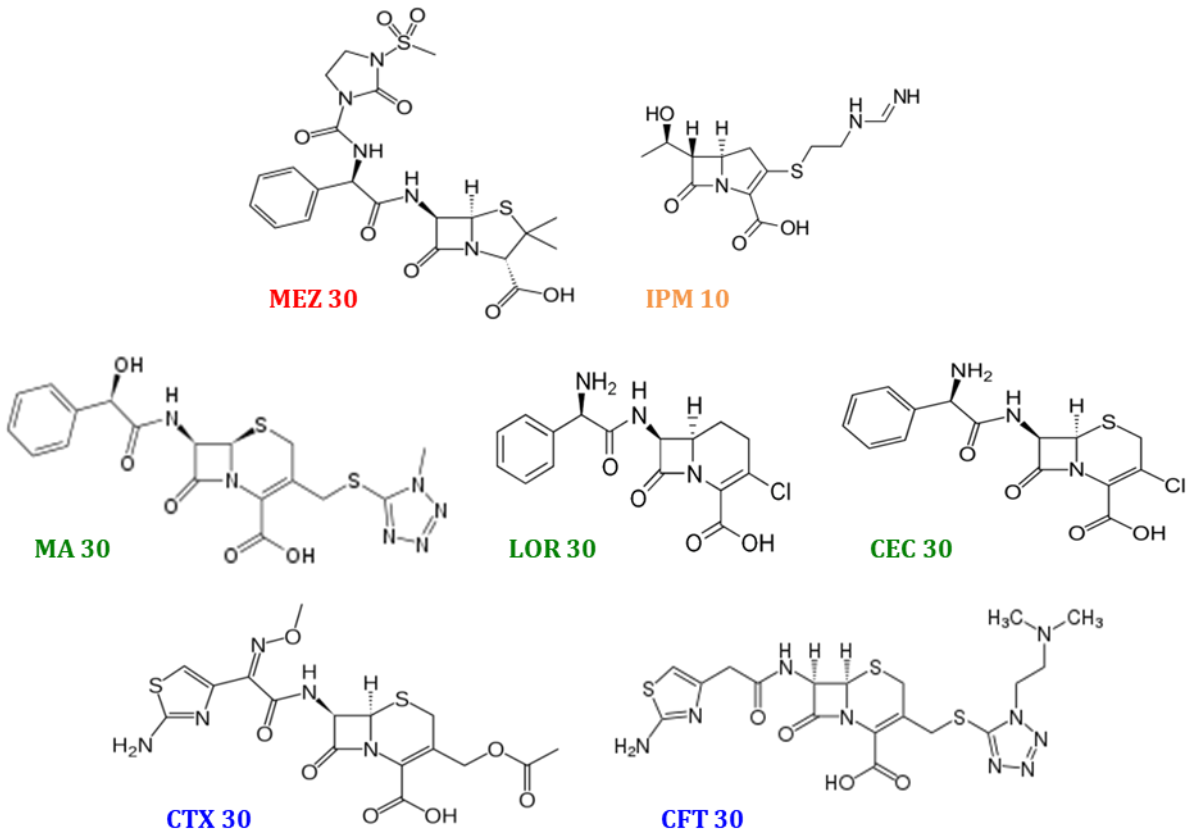
Zhang, C., P. L. Freddolino and Y. Zhang (2017). "COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information." Nucleic Acids Res **45**(W1): W291-W299.

Zhaxybayeva, O., K. S. Swithers, P. Lapierre, G. P. Fournier, D. M. Bickhart, R. T. DeBoy, K. E. Nelson, C. L. Nesbo, W. F. Doolittle, J. P. Gogarten and K. M. Noll (2009). "On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales." Proc Natl Acad Sci U S A **106**(14): 5865-5870.

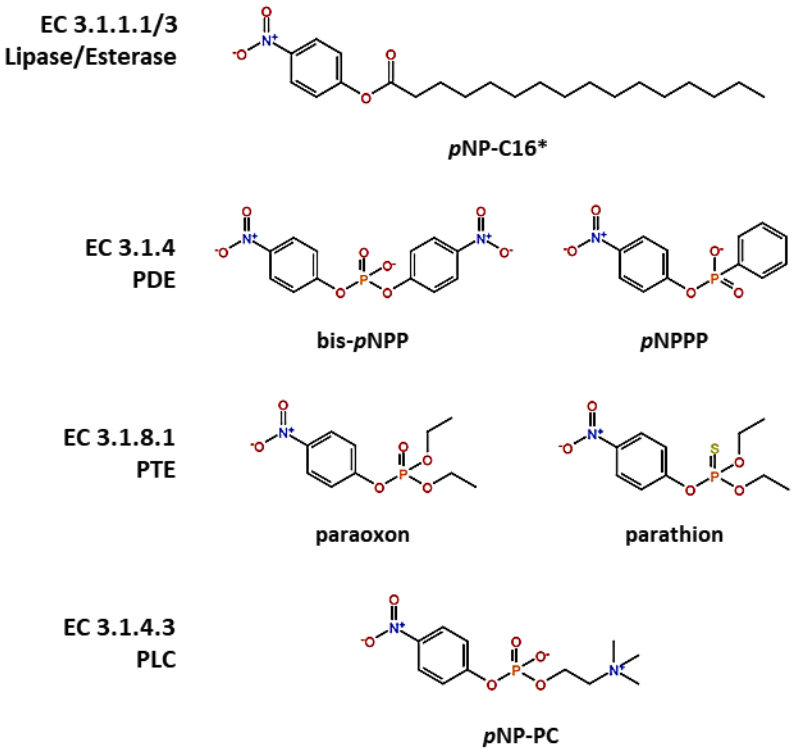
Zheng, X., N. Feng, D. Li, X. Dong and J. Li (2017). "New molecular insights into an archaeal RNase J reveal a conserved processive exoribonucleolysis mechanism of the RNase J family." Mol Microbiol **106**(3): 351-366.

SUPPLEMENTARY DATA

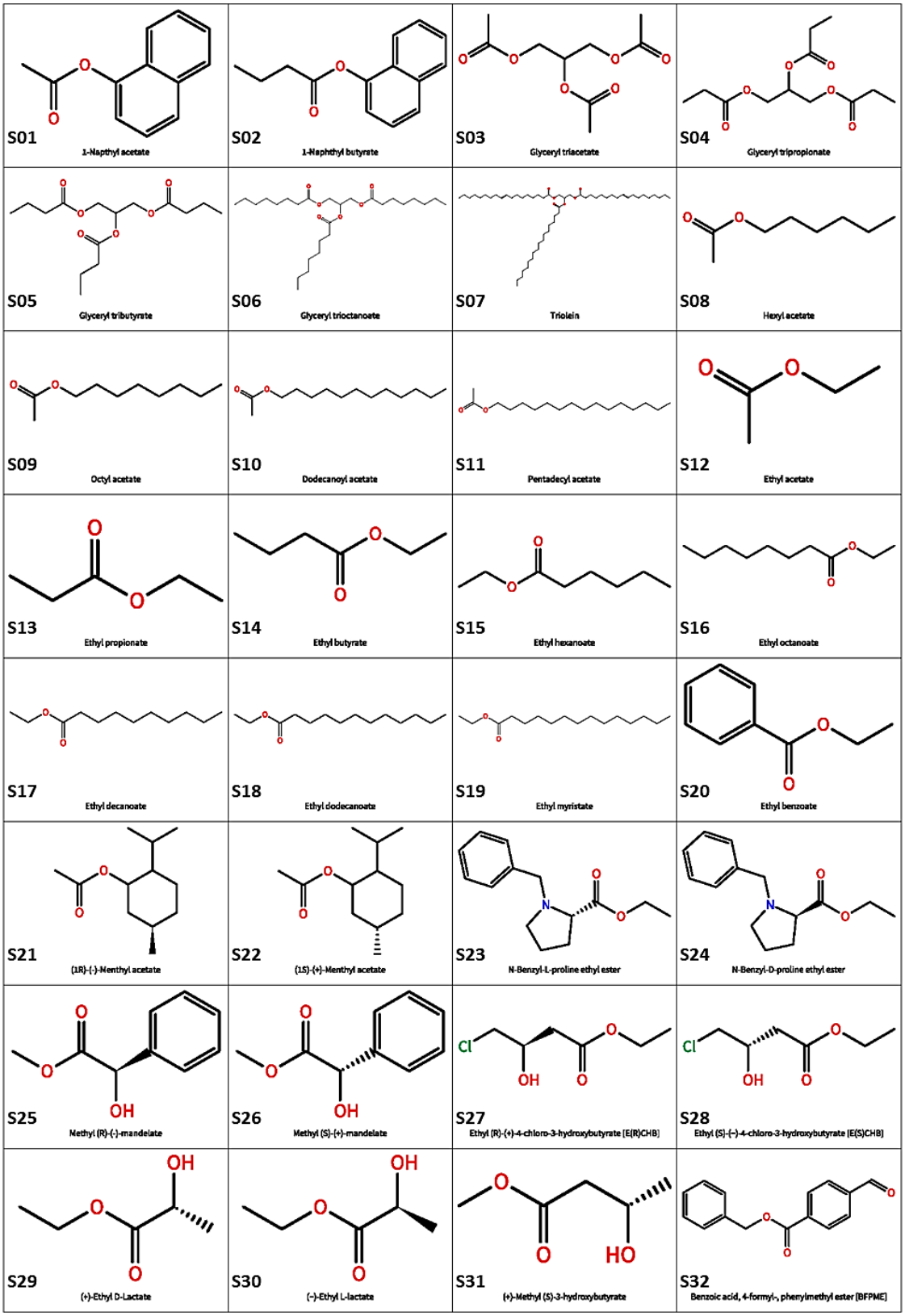
Supplementary data



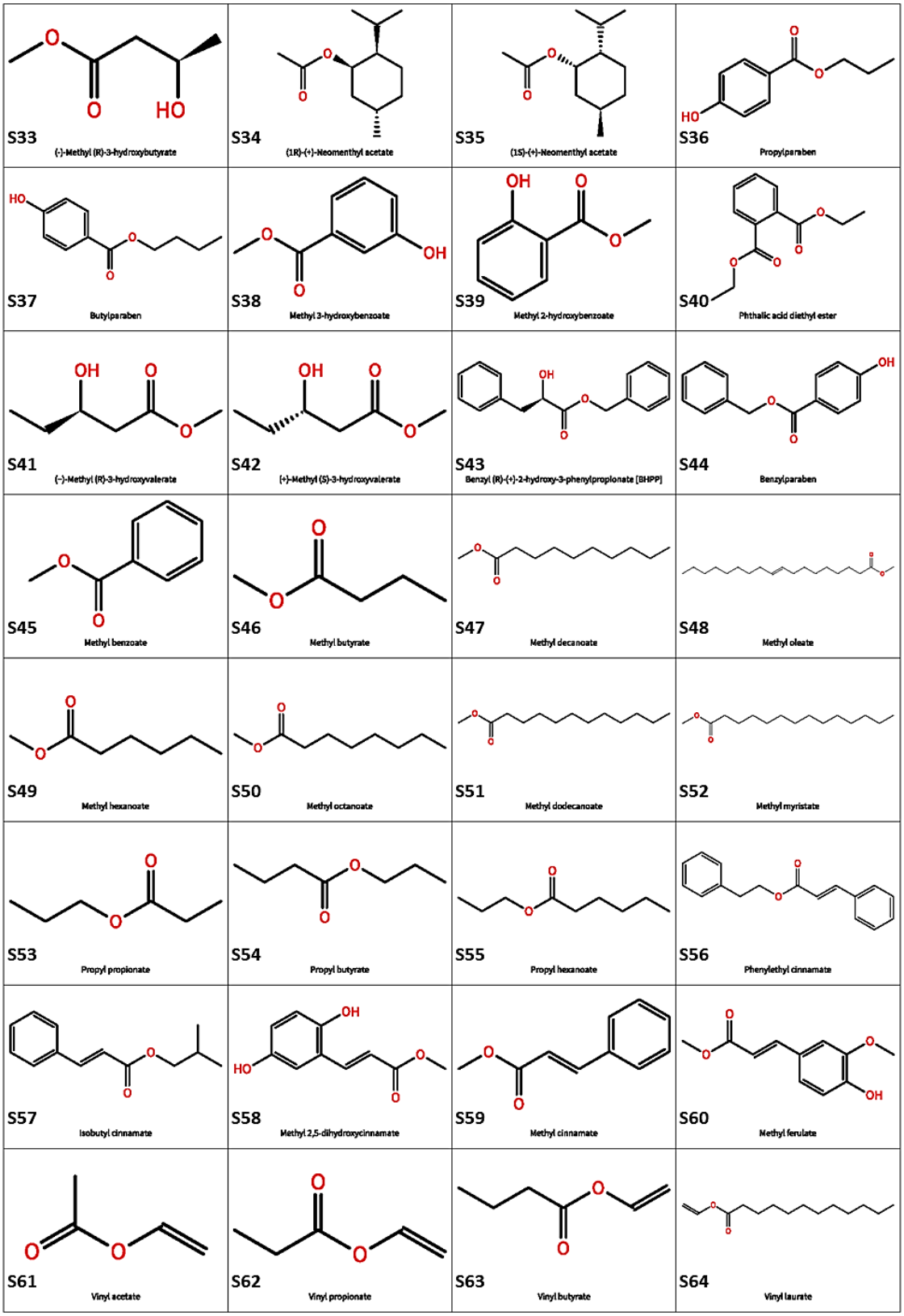
Supplementary Figure 1: Chemical structures of β-lactam antibiotics.
Mezlocillin 30 µg (MEZ 30), imipenem 10 µg (IPM 10), cefamandole 30 µg (MA 30), loracarbef 30 µg (LOR 30), cefaclor 30 µg (CEC 30), cefotaxime 30 µg (CTX 30) and cefotiam 30 µg (CFT 30). Penams are highlighted in red, carbapenems in orange, cephalosporines from the second generation in green and cephalosporines from the third generation in blue.



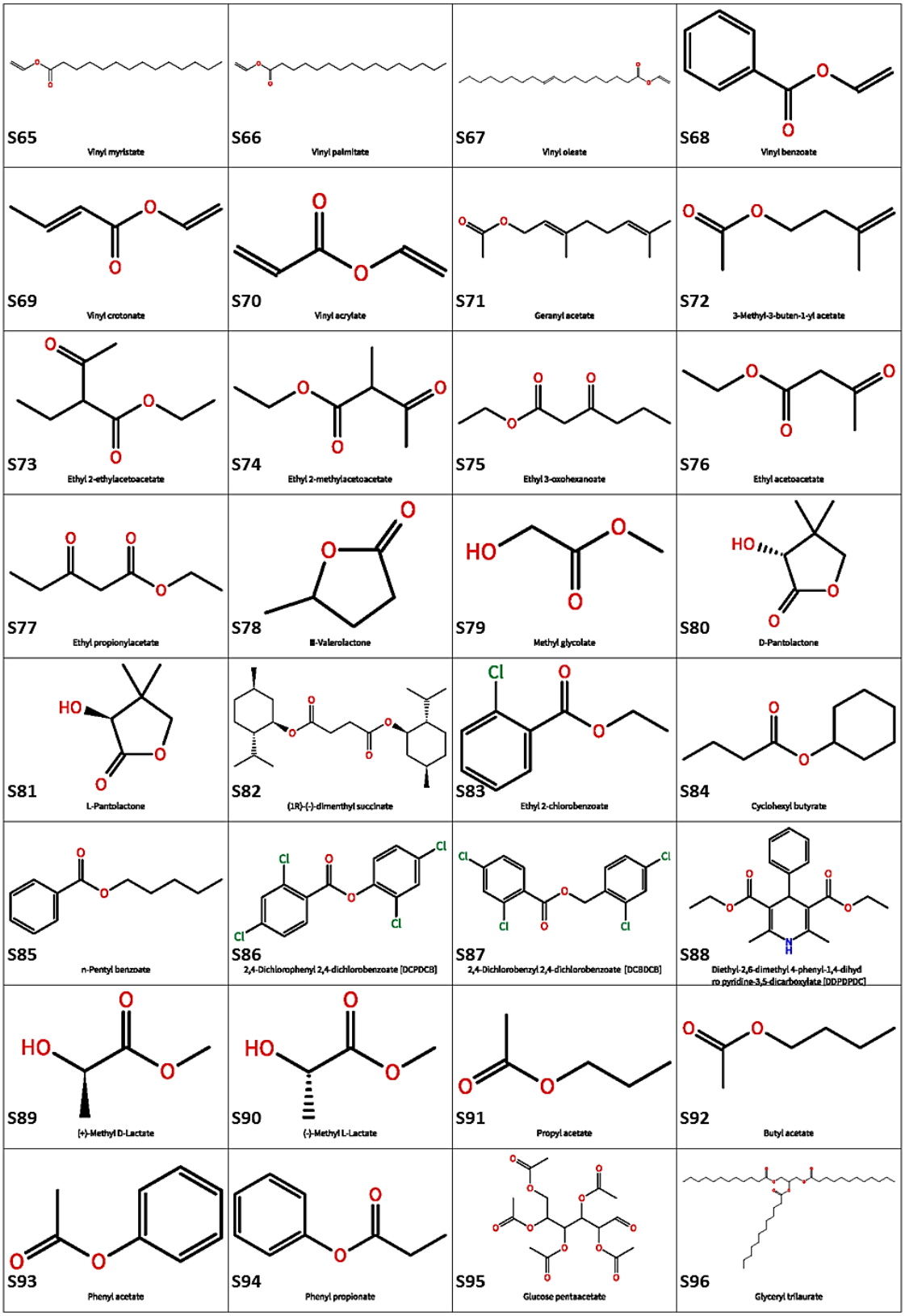
Supplementary Figure 2: *p*NP-substituted substrates degraded by Igni18.
*Substrates with acyl chains of C2, C4, C8, C12, C14 and C16 and 18 were used.



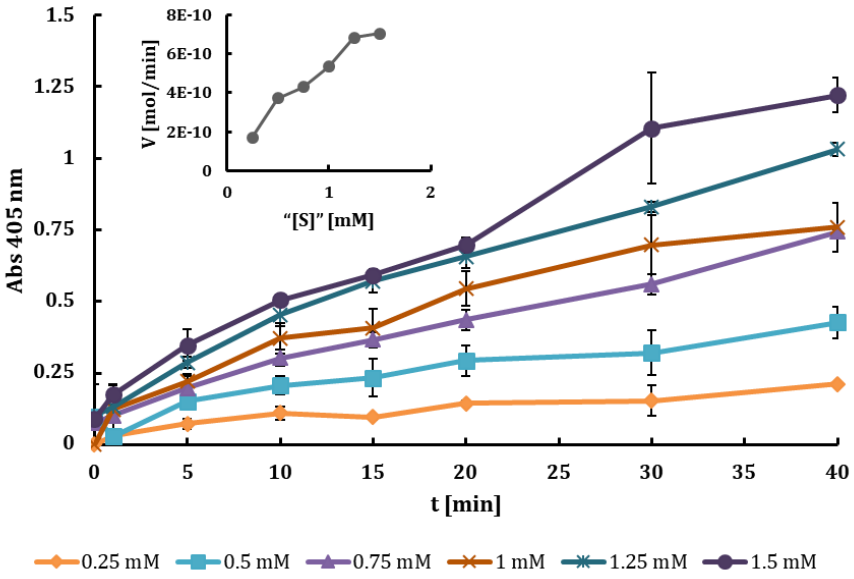
Supplementary Figure 3: Ester collection for "enzyme fingerprinting".



Supplementary Figure 3, cont.



Supplementary Figure 3, cont.



Supplementary Figure 4: Michaelis-Menten kinetics of Igni18 on bis-*p*NPP.
Experiments were conducted at 90 °C. Data represent mean values of triplicates and bars indicate standard error.

6HRG	0.00									
3X2Z	0.88	0.00								
4QN9	1.56	1.41	0.00							
3BV6	1.46	1.61	1.86	0.00						
2WYM	1.54	1.59	1.89	0.32	0.00					
6BRM	1.45	1.43	1.62	1.81	1.76	0.00				
5HAA	1.60	1.61	1.84	1.83	1.78	1.80	0.00			
4JO0	1.70	1.56	1.56	2.06	1.88	1.92	1.96	0.00		
2XR1	1.59	1.60	1.72	1.90	1.79	1.75	1.44	1.72	0.00	
4GCW	1.89	1.60	1.76	1.79	1.63	1.90	2.01	1.76	1.67	0.00
	6HRG	**3X2Z**	**4QN9**	**3BV6**	**2WYM**	**6BRM**	**5HAA**	**4JO0**	**2XR1**	**4GCW**
Avg. Dist.:	1.37	1.33	1.52	1.46	1.42	1.54	1.59	1.61	1.52	1.60

Supplementary Figure 5: Evolutionary divergence of 10 MβLs.
Estimates of evolutionary divergence between the non-ambiguous positions of a structural alignment including all structures used for PVR description. Depicted are the number of amino acid substitutions per site and the average distance of each structure towards the rest of them. Structures are sorted by overall structural similarity to Igni18.

Supplementary Table 1: Primers used in this study (Pt.2).

No.	Gene_name	Tann [°C]	Size [bp]	Restr_enzyme	Forward Primer	Tm [°C]	Reverse Primer	Tm [°C]
1	1_lipase_10c_11	60	687	NdeI / HindIII	CATATGCGCTACACCTCCCTTG	65	AAGCTTCTTCGCCACAGCCGCCTTC	73
2	1_lipase_25c_2	59	687	NdeI / HindIII	CATATGCGCGTCAAGTGCGT	66	AAGCTTCAGGGCTTCCTTGAC	64
3	1_esterase_1c_18	53	1491	NdeI / HindIII	CATATGAAAAAACTACTTTGCTTAGTG	59	AAGCTTTTGTTTTTTGATTGGAG	58
4	1_esterase_3c_12	53	2253	NdeI / HindIII	CATATGAATATCAACACATATTCTGC	58	AAGCTTGCACCAGCGGTCGAG	69
5	1_esterase_13c_9	64	1149	NdeI / HindIII	CATATGAAAAGACTGTCAATTATTTCCGCCATAGCAC	69	AAGCTTGAACAGCTGCGGGGCGAAGC	75
6	1_esterase_25c_9	58	1686	NdeI / HindIII	CATATGAAGGTATTGCATGAATGTGTCG	63	AAGCTTTTCGGCGGGATGACGAAC	64
7	1_esterase_30c_5	51	807	NdeI / XhoI	CATATGGGACTGTTTGTGGGAG	61	CTCGAGGTAAATCACCATTTCGTCAG	56
8	1_abhydrolase_14c_19	51	912	NdeI / XhoI	CATATGAATACAACAGAAATCGTTAC	56	CTCGAGTTTTATGCAAAATCTCTCTAGTAC	56
9	1_abhydrolase_32c_7	70	954	NdeI / HindIII	CATATGAGCAAGCAACAACGCGCCCGGATC	75	AAGCTTCGACCCGGCCGAGCGGCTCC	76
10	1_abhydrolase_32c_9	68	3642	NdeI / XhoI	CATATGAACCGCATCGCAATCCCGTTCGAC	73	CTCGAGCGCCCTCCCCGCCCGGGTGC	79
11	1_abhydrolase_33c_4	60	966	NdeI / HindIII	CATATGGCGTCACTTGCTGCAAA	65	AAGCTTTGCAACAGCCTCCTGCGTGG	69
12	1_abhydrolase_43c_3	57	909	NdeI / HindIII	CATATGACACGGTCCATGATTGG	62	AAGCTTTCGCCCCTCCAGGAAC	62
13	2_lipase_2c_7	52	630	NheI / XhoI	GCTAGCATGAATAAACTTGCTGAAATATTTAAG	57	CTCGAGATACAAAGCTTCATCATCGTAATG	59
14	2_esterase_26c_4	50	1152	NdeI / HindIII	CATATGAAAAGACTGTCAATTATTTCC	57	AAGCTTTTTGAACAGCTGCGG	55
15	2_esterase_21c_23	52	2181	NdeI / HindIII	CATATGCAGTTCATTATGGATGAG	57	AAGCTTGAATTTTGGTTTCATGTTCATATTG	57
16	2_esterase_21c_16	55	1368	NdeI / XhoI	CATATGAAAATAGTTTATGCATTTCTTGCCC	60	CTCGAGTCGCAACAGCCTTCTGACC	64
17	2_esterase_13c_22	56	1569	NdeI / XhoI	CATATGAAAAATTACCTTTTCCCCCTG	61	CTCGAGTTTCGCCCATTCCTTGCC	62
18	2_esterase_8c_8	55	1152	NheI / HindIII	GCTAGCATGGCGGCAATTCTTGAAC	61	AAGCTTCGCGTCTTCTTGCAGATG	60
19	2_esterase_6c_5	53	807	NdeI / XhoI	CATATGGGACTGTTTGTGGGAG	61	CTCGAGGTAAATCACCATTTCGTCAGG	58
20	2_esterase_3c_18	51	2253	NdeI / HindIII	CATATGAATATCAACACATATTCTGC	56	AAGCTTTTTGCACCAGCGGTCG	62
21	2_abhydrolase_31c_3	53	2472	NdeI / HindIII	CATATGAAGAGTAAACTTTTCCTTCC	58	AAGCTTATCATTGTTCATCTGCGGG	59
22	2_abhydrolase_21c_9	50	1767	NdeI / HindIII	CATATGATTTGGAAGGAAACAAC	56	AAGCTTTTTGTTGAATCCTTTAGCAG	55
23	2_abhydrolase_18c_23	52	831	NheI / HindIII	GCTAGCATGAAGTCTAAATTATTGGACACAG	59	AAGCTTCTGAATACGACAACCAACC	57
24	2_abhydrolase_4c_6	53	951	NdeI / XhoI	CATATGAAAGGAGATAATGATGCTTTAAC	59	CTCGAGACACCTGCAGAAAGCC	58
25	3_lipase_1c_6	53	1173	NdeI / HindIII	CATATGCGTTTATACAAATTGATATTAG	56	AAGCTTATGCCGGCCATGC	57
26	3_lipase_1c_23	60	582	NdeI / HindIII	CATATGCTCGGTATTGTATTGCTGCTGCCC	67	AAGCTTCAGCATCCCCTTCAACTTCTGTAG	64
27	3_lipase_3c_12	56	2352	NdeI / HindIII	CATATGAAAAATAAGTCGAACCAGCTCCTA	62	AAGCTTGTTTGCAGGCAACTCAATG	59
28	3_lipase_7c_26	54	1212	NheI / HindIII	GCTAGCATGCACCAACCTCTTGGC	62	AAGCTTATGATGCTTAATCCAACCG	57
29	3_lipase_51c_6	56	1332	NheI / HindIII	GCTAGCATGCCGGTCATCGTCATTC	62	AAGCTTTTTGTTTTTCGAGACGGTTTTG	60
30	3_abhydrolase_1c_37	54	705	NdeI / XhoI	CATATGGCTAACAGAATCTCAACGC	58	CTCGAGTTTGCGGATTTGATTGTTGG	59
31	3_abhydrolase_2c_44	56	759	NdeI / HindIII	CATATGCCGTTCCTGACCGTTCC	61	AAGCTTGGCCTCGCCCAAGAAC	62
32	3_abhydrolase_3c_14	54	579	NdeI / HindIII	CATATGGCTGCACTGCTGTACATC	60	AAGCTTGTGTTTAAAAAATGTCATGCAATC	58
33	3_abhydrolase_13c_2	54	1089	NdeI / HindIII	CATATGTCGATCGTTTATTCGCTGTTAC	60	AAGCTTGACTGCGCCGGTTTC	59
34	3_abhydrolase_16c_31	55	915	NdeI / HindIII	CATATGAAAACTTGGAAATCCCTTGTCG	61	AAGCTTCCGCTTCAAGTGCTGATCAA	62
35	3_abhydrolase_19c_35	51	1503	NdeI / HindIII	CATATGCGCACATTTGAAATCTTGAC	57	AAGCTTTTGCGAATGATTGTATATCTT	55
36	3_abhydrolase_20c_2	51	879	NdeI / HindIII	CATATGCCCGCCCTGACTTC	56	AAGCTTGGCTATGAATGTCGTGAC	57
37	3_abhydrolase_20c_11	57	1917	NdeI / HindIII	CATATGATTTCCACCCCGAACCGAAC	64	AAGCTTCTGGCTCCCAGCTTCC	61
38	3_abhydrolase_22c_10	52	951	NdeI / HindIII	CATATGAAGGTTGTTATTCGGATCAG	57	AAGCTTCAGCGTTGCTTGCAG	57
39	3_abhydrolase_23c_15	51	828	NheI / HindIII	GCTAGCATGAAGTCTAAATTATTGGACAC	56	AAGCTTCTGAATACGACAACCAAC	55
40	3_abhydrolase_29c_14	65	855	NdeI / HindIII	CATATGAACTGGTGGTGGGCCGTCCTG	71	AAGCTTGCTGCGCAGGGCGCGATC	71
41	3_esterase_2c_12	55	945	NdeI / HindIII	CATATGCTTCGGTTGGGTCTTGTC	59	AAGCTTGCGCGCCCGGCTCTTG	69
42	3_esterase_4c_10	52	1146	NdeI / HindIII	CATATGAAAAGACTGTCAATTATTTCCG	56	AAGCTTTTTGAACAGCTGCGGGGC	65
43	3_esterase_4c_27	59	462	NdeI / HindIII	CATATGCTTCGGAAGCGACGAGATTCA	64	AAGCTTTGCTCCCGGAATGACGTTG	64
44	3_esterase_9c_7	59	2178	NdeI / HindIII	CATATGGTCTTGCGACGGCCTCTG	65	AAGCTTGGGTACGTCGTAGGCGTC	63
45	3_esterase_10c_9	57	774	NdeI / HindIII	CATATGAAGACCCTGACGACCGCAATC	66	AAGCTTTTTGAATATCTCGGCGAAGAAC	60
46	3_esterase_14c_25	53	2013	NdeI / XhoI	CATATGCCAACCCCCAAAAAACG	58	CTCGAGCTTCAAATACTTATCCATCCAG	56
47	3_esterase_15c_26	52	1470	NheI / HindIII	GCTAGCATGACTCATCAAATAGTAACG	55	AAGCTTTTCTCCTTTTGGAGGGAATAG	58
48	3_esterase_16c_29	53	186	NdeI / HindIII	CATATGCCGATCTGGGTCTTCC	57	AAGCTTTTGCCGCAACCAGTC	58
49	3_esterase_16c_36	55	390	NdeI / HindIII	CATATGATACGTCTTGTCACCATTGC	60	AAGCTTCCGGCCCTTCAACAAC	59
50	3_esterase_20c_28	60	801	NdeI / HindIII	CATATGCCGCATATCAATCGCAACGG	64	AAGCTTGGCCGCGACGCGGG	68
51	3_esterase_24c_11	55	978	NdeI / XhoI	CATATGCCGCTAACTTCATTTGATTACTC	59	CTCGAGAAGACCCAGTAAAAACTCGAAAAC	61
52	3_esterase_29c_9	61	474	NdeI / HindIII	CATATGCCCGCCTTCCAGGACTACTAC	65	AAGCTTGCCTGCAGCCTGGCCG	68
53	3_esterase_50c_1	54	720	NdeI / XhoI	CATATGACCATGATTACGCCAAGC	59	CTCGAGTAGACCCAGAATCTTCTTCAC	58
54	3_esterase_51c_6	56	1332	NheI / HindIII	GCTAGCATGCCGGTCATCGTCATTC	62	AAGCTTTTTGTTTTTCGAGACGGTTTTG	60
55	3_esterase_65c_2	52	1536	NdeI / HindIII	CATATGCACAAAAAAATTTTAGCGATTATC	56	AAGCTTTTGGTTATTTTTCAAATACTCTAAAATC	57
56	4_lipase_1c_22	52	1872	NdeI / HindIII	CATATGTCACTCAAGAAAAAACTCATACC	58	AAGCTTGAAGTCGACCGAAACG	56
57	4_lipase_23c_6	52	582	NdeI / HindIII	CATATGCTCGGTATTGTATTGCTGC	58	AAGCTTCAGCATCCCCTTCAAC	56
58	4_lipase_23c_23	52	1173	NdeI / HindIII	CATATGCGTTTATACAAATTGATATTAGCG	56	AAGCTTATGCCGGCCATGC	57
59	4_esterase_8c_37	59	1635	NdeI / HindIII	CATATGAAGGTCGAGAAAGGCATCAC	62	AAGCTTGGCGGCGGCGATCAC	65
60	4_esterase_9c_19	53	2184	NdeI / NotI	CATATGAAACGAATTCTAACTTTCTTCGC	59	GCGGCCGCGTTTGGGGGCTTTTCATC	58
61	4_esterase_9c_28	51	996	NdeI / NotI	CATATGACAAAGCTTTTAGAACTGTAC	56	GCGGCCGCTTTGCTACTTAGGAACTTCAA	56
62	4_esterase_22c_7	56	2187	NdeI / HindIII	CATATGGTCTTGCGACGGCCTCT	63	AAGCTTGGGTACGTCGTAGGCG	60
63	4_esterase_24c_4	52	1854	NdeI / HindIII	CATATGGGTGTGTACGACTACAAAAAC	59	AAGCTTGGCGATCACAATCCC	55
64	4_esterase_24c_3	55	1701	NdeI / HindIII	CATATGGGGCTGTTAGATTACAAAAATGG	59	AAGCTTTGCAAACACAAACCCGG	60
65	4_esterase_245c_1	55	489	NdeI / HindIII	CATATGTTTGATTATCTGCGCGATG	58	AAGCTTCTTCAGTAAATACTGCGCATGG	61
66	4_esterase_280c_1	50	486	NdeI / HindIII	CATATGTCTGCCGTACTGACC	55	AAGCTTACCTACCATATAGGTGGC	56
67	4_esterase_326c_1	56	213	NdeI / HindIII	CATATGGCCGGATTTATGATGACCC	60	AAGCTTTCCCAAAACTGCCGTACC	61
68	4_esterase_806c_1	53	405	NdeI / HindIII	CATATGCTGTTCGGGCAGAGTATTG	59	AAGCTTGTTCTGCGGATTCAACG	57
69	4_abhydrolase_6c_31	53	864	NdeI / HindIII	CATATGAACAGGCAGGCTTGG	57	AAGCTTCGCTCGCACGCCTTC	63
70	4_abhydrolase_7c_19	51	888	NdeI / HindIII	CATATGATGTCATCTTCTTTTGAAAGC	56	AAGCTTCCCCAAGTGCGATTTAAGG	60
71	4_abhydrolase_8c_31	55	705	NdeI / HindIII	CATATGCCCCACACCTTCGTCC	60	AAGCTTGGCTGCCTCTAAGAGAATC	59
72	4_abhydrolase_10c_19	57	927	NdeI / HindIII	CATATGGCTTTCCTGGCCATAGGG	61	AAGCTTGCAGCCGCGAGGAATTG	63
73	4_abhydrolase_11c_22	52	888	NdeI / HindIII	CATATGAAGAAAGCCTCTTTACGTTG	57	AAGCTTAGGAATATTCAGCAAGTGG	56
74	4_abhydrolase_15c_28	56	1401	NdeI / HindIII	CATATGAAACTTGCGAAGATCGTGATG	61	AAGCTTCTGCGCTGCAGCC	59
75	4_abhydrolase_18c_8	56	696	NdeI / HindIII	CATATGACCACATTCGTATTGGTGC	60	AAGCTTCGCTAGTCCAGCAACCAG	62
76	4_abhydrolase_18c_13	55	945	NdeI / HindIII	CATATGCAGATCGATCCCGAATGC	59	AAGCTTAAAATTCGCCTTGATCACGG	61
77	4_abhydrolase_20c_20	52	912	NdeI / HindIII	CATATGGCCAGTCCACAGCTAC	58	AAGCTTAACAGTGGCTCCTGTG	57
78	4_abhydrolase_31c_8	61	873	NdeI / HindIII	CATATGGATTCGCTGCCGCTGCTC	66	AAGCTTGCGGTGCGCGGCCAG	69
79	4_abhydrolase_33c_10	47	762	NdeI / XhoI	CATATGGCTAACAGAATCTCAAC	51	CTCGAGTTTAGGTGAGACTATAGAATAC	53
80	4_abhydrolase_35c_4	57	951	NdeI / HindIII	CATATGCTGGTGGCGGTGACG	61	AAGCTTTGGCGTGCCGCCG	64

Supplementary Table 2: Crystallization conditions.
This table is extracted from Kobus *et al.*, 2019.

Method	Sitting drop vapor diffusion
Plate type	MRC3 Swissci
Temperature (K)	293
Protein concentration (mg/mL)	20
Buffer composition of protein solution	0.1 M potassium phosphate pH 7
Composition of reservoir solution	0.3 M $Mg(NO_3)_2 \cdot 6\ H_2O$ 0.1 M Tris pH 8; 22 % (w/v) PEG 8000
Volume and ratio of drop	200 nL, 1:1
Volume of reservoir (µL)	40

Supplementary Table 3: Data collection and refination statistics.
Statistics for the highest-resolution shell are shown in parentheses. This table is extracted from Pérez-García *et al.*, submitted.

Wavelength	0.9677
Resolution range	31.32 - 2.3 (2.382 - 2.3)
Space group	R 3 2: H
Unit cell	67.42 67.42 253.77 90 90 120
Total reflections	70412 (7119)
Unique reflections	10293 (1008)
Multiplicity	6.8 (7.1)
Completeness (%)	99.78 (100.00)
Mean I/sigma(I)	12.60 (3.37)
Wilson B-factor	32.46
R-merge	0.1004 (0.5445)
R-meas	0.1084 (0.5872)
R-pim	0.03987 (0.2152)
CC1/2	0.998 (0.885)
CC*	1 (0.969)
Reflections used in refinement	10282 (1008)
Reflections used for R-free	500 (54)
R-work	0.1730 (0.1866)
R-free	0.2108 (0.2179)
CC (work)	0.941 (0.935)
CC (free)	0.950 (0.906)
Number of non-hydrogen atoms	1908
macromolecules	1803
ligands	8
solvent	97
Protein residues	233
RMS (bonds)	0.008
RMS (angles)	1.26
Ramachandran favored (%)	93.51
Ramachandran allowed (%)	5.62
Ramachandran outliers (%)	0.87
Rotamer outliers (%)	0.52
Clashscore	3.6
Average B-factor	33.9
macromolecules	33.68
ligands	38.59
solvent	37.61

Supplementary Table 4: Structural search against the PDB.
Results obtained with mTM-align (Dong *et al.* 2018). The first hundred hits are displayed. Only one hit per monomer of a multimeric enzyme is shown.

No.	subject	TM-score	RMSD	Sequence id.	Subject length	Aligned length	Description
1	3x2zA	0.8963	1.28	0.413	227	218	CRYSTAL STRUCTURE OF METALLO-BETA-LACTAMASE IN COMPLEX WITH NICKEL FROM THERMOTOGA MARITIMA
2	3x2xB	0.8955	1.29	0.408	227	218	CRYSTAL STRUCTURE OF METALLO-BETA-LACTAMASE H48A FROM THERMOTOGA MARITIMA
7	4qn9B	0.88	2.12	0.206	322	228	STRUCTURE OF HUMAN NAPE-PLD
8	3x2yA	0.8794	1.9	0.408	227	218	CRYSTAL STRUCTURE OF METALLO-BETA-LACTAMASE H8A FROM THERMOTOGA MARITIMA
13	3bv6A	0.8418	2.69	0.211	353	228	CRYSTAL STRUCTURE OF UNCHARACTERIZED METALLO PROTEIN FROM VIBRIO CHOLERAE WITH BETA-LACTAMASE LIKE FOLD
21	2wymA	0.8334	2.55	0.194	303	222	STRUCTURE OF A METALLO-B-LACTAMASE
26	2wylA	0.816	2.71	0.191	296	220	APO STRUCTURE OF A METALLO-B-LACTAMASE
31	6brmB	0.8049	2.92	0.197	263	223	THE CRYSTAL STRUCTURE OF ISOTHIOCYANATE HYDROLASE FROM DELIA RADICUM GUT BACTERIA
32	5habB	0.8046	2.92	0.186	463	221	CRYSTAL STRUCTURE OF MPY-RNASE J (MUTANT H84A) AN ARCHAEAL RNASE J FROM METHANOLOBUS PSYCHROPHILUS R15 COMPLEX WITH RNA
38	5haaB	0.8006	2.98	0.19	460	221	CRYSTAL STRUCTURE OF MPY-RNASE J AN ARCHAEAL RNASE J FROM METHANOLOBUS PSYCHROPHILUS R15
39	4jo0A	0.7996	3.1	0.162	522	222	CRYSTAL STRUCTURE OF CMLA A DIIRON BETA-HYDROXYLASE FROM STREPTOMYCES VENEZUELAE
41	5ws2B	0.7994	2.99	0.19	463	221	CRYSTAL STRUCTURE OF MPY-RNASE J (MUTANT S247A) AN ARCHAEAL RNASE J FROM METHANOLOBUS PSYCHROPHILUS R15 COMPLEX WITH RNA
45	5kikA	0.7968	3.05	0.158	521	221	CMLA BETA-HYDROXYLASE IN CHEMICALLY REDUCED DIFERROUS STATE
48	5a0vA	0.7932	2.95	0.164	454	219	CATALYSIS AND 5 END SENSING BY RIBONUCLEASE RNASE J OF THE METALLO-BETA-LACTAMASE FAMILY
49	5a0tA	0.793	2.94	0.164	454	219	CATALYSIS AND 5 END SENSING BY RIBONUCLEASE RNASE J OF THE METALLO-BETA-LACTAMASE FAMILY
51	2xr1A	0.792	2.89	0.181	605	216	DIMERIC ARCHAEAL CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR WITH N-TERMINAL KH DOMAINS (KH-CPSF) FROM METHANOSARCINA MAZEI
54	3af5A	0.7912	2.87	0.148	638	216	THE CRYSTAL STRUCTURE OF AN ARCHAEAL CPSF SUBUNIT PH1404 FROM PYROCOCCUS HORIKOSHII
55	2ycbA	0.791	2.94	0.167	634	216	STRUCTURE OF THE ARCHAEAL BETA-CASP PROTEIN WITH N-TERMINAL KH DOMAINS FROM METHANOTHERMOBACTER THERMAUTOTROPHICUS
58	2i7tA	0.7891	2.91	0.191	404	215	STRUCTURE OF HUMAN CPSF-73
59	6i1dA	0.7883	2.91	0.181	458	215	STRUCTURE OF THE YSH1-MPE1 NUCLEASE COMPLEX FROM S.CEREVISIAE
62	4xwwA	0.7864	2.98	0.169	545	219	CRYSTAL STRUCTURE OF RNASE J COMPLEXED WITH RNA
64	3rpcA	0.7858	2.99	0.189	252	217	THE CRYSTAL STRUCTURE OF A POSSIBLE METAL-DEPENDENT HYDROLASE FROM VEILLONELLA PARVULA DSM 2008
68	3zq4A	0.7843	2.97	0.183	550	218	UNUSUAL DUAL ENDO- AND EXO-NUCLEASE ACTIVITY IN THE DEGRADOSOME EXPLAINED BY CRYSTAL STRUCTURE ANALYSIS OF RNASE J1
71	4ojxA	0.7836	3.27	0.152	363	223	CRYSTAL STRUCTURE OF YEAST PHOSPHODIESTERASE-1 IN COMPLEX WITH GMP
72	6j4nA	0.7835	2.92	0.125	372	216	STRUCTURE OF PAPUA NEW GUINEA MBL-1(PNGM-1) NATIVE
74	4ojvA	0.7831	3.25	0.152	363	223	CRYSTAL STRUCTURE OF UNLIGANDED YEAST PDE1
83	4gcwA	0.7757	2.82	0.136	307	213	CRYSTAL STRUCTURE OF RNASE Z IN COMPLEX WITH PRECURSOR TRNA(THR)
84	3t3nA	0.7746	3.11	0.192	550	219	MOLECULAR BASIS FOR THE RECOGNITION AND CLEAVAGE OF RNA (UUCCGU) BY THE BIFUNCTIONAL 5 -3 EXO/ENDORIBONUCLEASE RNASE J
85	2fk6A	0.7734	2.97	0.126	307	214	CRYSTAL STRUCTURE OF RNASE Z/TRNA(THR) COMPLEX
88	2cbnA	0.7644	3.02	0.136	306	213	CRYSTAL STRUCTURE OF ZIPD FROM ESCHERICHIA COLI
89	4z67A	0.7641	3.14	0.127	309	220	THE 1.5-ANGSTROM CRYSTAL STRUCTURE OF MN(2+)-BOUND PQQB FROM PSEUDOMONAS PUTIDA
91	4z5zA	0.7624	3.16	0.127	308	220	THE 2.5-ANGSTROM CRYSTAL STRUCTURE OF MG(2+)-BOUND PQQB FROM PSEUDOMONAS PUTIDA
92	1y44A	0.7616	2.74	0.139	265	209	CRYSTAL STRUCTURE OF RNASE Z
93	6e13A	0.7615	3.2	0.123	308	220	PSEUDOMONAS PUTIDA PQQB WITH A NON-PHYSIOLOGICAL ZINC AT THE ACTIVE SITE BINDS THE SUBSTRATE MIMIC
95	3bk1A	0.7613	3.16	0.194	537	217	CRYSTAL STRUCTURE ANALYSIS OF RNASE J
96	5mtzA	0.7604	3.26	0.111	750	216	CRYSTAL STRUCTURE OF A LONG FORM RNASE Z FROM YEAST
97	4z60A	0.7593	3.22	0.123	308	220	THE 2.5-ANGSTROM OF CRYSTAL STRUCTURE OF ZN(2+)-BOUND PQQB FROM PSEUDOMONAS PUTIDA
98	4z6xA	0.7584	3.21	0.136	309	220	THE 1.68-ANGSTROM CRYSTAL STRUCTURE OF ACITIVE-SITE METAL-FREE PQQB FROM PSEUDOMONAS PUTIDA

Supplementary Table 5: Definition of PVRs for a set of MβLs.
The start and end amino acid positions of each PVR are shown for each MβL analyzed. The RMSD values for the structural alignment at each position are also given. This table is extracted from Pérez-García et al., submitted.

ENZYME	PDB	PVR1			PVR2			PVR3			PVR4			PVR5			PVR6			PVR7			PVR8			PVR9			PVR10		
		start	end	length	start	end	length	start	end	length	start	end	length	start	end	length	start	end	length	start	end	length	start	end	length	start	end	length	start	end	length
Igni18	6HRG	1	K 5	5	H 11	A 13	3	W 28	T 47	20	I 77	G 91	15	T 117	P 126	10	F 132	L 138	7	L 147	P 160	14	G 168	M 173	6	Y 195	V 220	26	T 230	K 234	5
MβL	3X2Z	1	K 2	2	H 8	V 10	3	F 24	I 41	18	N 71	T 85	15	V 107	P 124	18	T 130	K 135	6	L 144	V 156	13	G 164	M 169	6	Y 191	C 215	25	E 225	L 226	2
ITCase	6BRM	1	K 2	2	N 8	T 10	3	M 25	V 60	36	Q 90	I 106	17	Q 127	A 146	20	T 152	T 158	7	W 167	P 180	14	G 188	M 200	13	M 222	I 248	27	A 258	TRP 262	5
β-hydroxylase	4JO0	1	V 252	252	H 258	C 260	3	V 275	I 298	24	P 329	V 353	25	L 375	K 386	12	R 392	S 397	6	N 406	V 422	17	E 430	S 461	32	M 483	A 521	39	R 531	LEU 532	2
NAPE-PLD	4QN9	1	R 126	126	H 132	T 134	3	I 149	P 177	29	P 212	V 227	16	Q 252	W 267	16	L 273	R 278	6	Y 287	F 301	15	G 309	V 322	14	W 344	F 375	32	Y 385	ASP 389	5
KH-CPSF1	2XR1	1	R 182	182	G 188	C 196	9	G 211	I 235	25	T 266	T 305	40	G 328	S 333	6	H 339	N 346	8	Y 355	V 371	17	T 379	S 581	203	G 605	T 625	21	R 635	LEU 637	3
RNASE J	5HAA	1	G 5	5	G 11	M 19	9	G 34	V 75	42	T 105	I 128	24	Q 151	T 157	7	H 163	A 168	6	F 177	V 200	24	T 208	A 389	182	G 411	L 434	24	Y 444	GLU 447	4
tRNase Z	4GCW	1	E 2	2	T 8	S 21	14	G 39	I 56	18	P 93	I 117	25	I 139	A 144	6	Q 150	S 205	56	V 214	C 225	12	T 233	S 248	16	I 270	S 293	24	N 303	GLY 307	5
UlaG	2WYM	1	A 39	39	C 45	G 47	3	W 63	I 110	48	P 144	C 161	18	F 183	A 209	27	K 215	S 220	6	Y 229	I 242	14	G 250	M 260	11	H 282	P 312	31	T 322	PHE 338	17
Unknown function	3BV6	1	S 39	39	C 45	G 47	3	W 63	I 110	48	P 144	C 161	18	D 184	A 210	27	E 216	S 221	6	Y 230	I 243	14	G 251	M 261	11	H 283	P 313	31	T 323	LEU 355	33
RMSD [Å]		---	**1.64**		**1.651**	**1.251**		**1.907**	**1.801**		**1.498**	**1.918**		**1.572**	**1.43**		**1.659**	**1.417**		**1.738**	**1.296**		**1.979**	**1.9**		**1.574**	**1.959**		**1.356**	---	

Supplementary Table 6: BLASTn searches against the nr-database.
Only the Sanger reads that could be linked to a contig are displayed.

Sanger read	LipH_f_#	Contig matched	BLASTn of Sanger read	BLASTn of contig	Observations
Lip041_for	1	LipH_f_1_38c	---	---	---
Lip041_rev	1	LipH_f_1_25c	---	---	---
Lip051_for	2	LipH_f_2_21c	---	---	---
Lip051_rev	2	LipH_f_2_21c	---	---	---
Lip061_for	2	LipH_f_2_6c	---	---	Also matches LipH_f_1_26c completely
Lip061_rev	2	LipH_f_2_6c	---	---	Also matches LipH_f_1_30c completely
Lip086_for	3	LipH_f_3_4c	---	*Prevotella ruminicola/dentivola* (600 bp)	Also matches LipH_f_2_46c and LipH_f_1_13c completely
Lip086_rev	3	LipH_f_3_4c	---	*Prevotella ruminicola/dentivola* (600 bp)	Also matches LipH_f_2_29c completely
Lip087_for	3	LipH_f_3_4c	---	*Prevotella ruminicola/dentivola* (600 bp)	Also matches LipH_f_2_46c and LipH_f_1_13c completely
Lip087_rev	3	LipH_f_3_4c	---	*Prevotella ruminicola/dentivola* (600 bp)	Also matches LipH_f_2_29c completely
Lip088_for	3	LipH_f_3_23c	---	Uncultured bacterium pUR16A2 genomic sequence (1.5 kbp)	Also matches LipH_f_2_18c completely
Lip088_rev	3	LipH_f_3_23c	---	Uncultured bacterium pUR16A2 genomic sequence (1.5 kbp)	Also matches LipH_f_2_18c completely
Lip090_for	3	LipH_f_3_4c	---	*Prevotella ruminicola/dentivola* (600 bp)	Also matches LipH_f_2_46c
Lip090_rev	3	LipH_f_3_4c	---	*Prevotella ruminicola/dentivola* (600 bp)	Also matches LipH_f_2_46c
Lip091_for	3	LipH_f_3_4c	---	*Prevotella ruminicola/dentivola* (600 bp)	Also matches LipH_f_2_46c
Lip091_rev	3	LipH_f_3_4c	---	*Prevotella ruminicola/dentivola* (600 bp)	Also matches LipH_f_2_46c
Lip105_for	3	LipH_f_3_13c	---	Uncultured thaumarchaeote 4-hydroxybutyryl-CoA dehydratase (4-hbd) gene, partial cds	---
Lip105_rev	3	LipH_f_3_13c	---	Uncultured thaumarchaeote 4-hydroxybutyryl-CoA dehydratase (4-hbd) gene, partial cds	---
Lip107_for	3	LipH_f_3_51c	---	---	---
Lip107_rev	3	LipH_f_3_43c	---	---	---
Lip108_for	3	LipH_f_3_2c	---	Same short region for various organisms (*Hyphomicrobium*, *Methylobacterium*, ...)	---
Lip108_rev	3	LipH_f_3_2c	---	Same short region for various organisms (*Hyphomicrobium*, *Methylobacterium*, ...)	---
Lip109_for	3	LipH_f_3_16c	---	---	---
Lip109_rev	3	LipH_f_3_16c	---	---	---
Lip111_for	3	LipH_f_3_15c	*Bacillus sp.*	*Bacillus sp.*	---
Lip111_rev	3	LipH_f_3_15c	*Bacillus sp.*	*Bacillus sp.*	---
Lip112_for	3	LipH_f_3_7c	---	*Sideroxydans lithotrophicus* ES-1, complete genome (650 bp)	---
Lip112_rev	3	LipH_f_3_7c	---	*Sideroxydans lithotrophicus* ES-1, complete genome (650 bp)	---
Lip113_for	3	LipH_f_3_24c	---	---	Also matches LipH_f_4_32c completely
Lip113_rev	3	LipH_f_3_24c	---	---	Also matches LipH_f_4_38c completely
Lip114_for	3	LipH_f_3_19c	---	Only some short sequences, various origins	---
Lip114_rev	3	LipH_f_3_19c	---	Only some short sequences, various origins	---
Lip120_for	3	LipH_f_3_20c	---	*Rhodopseudomonas, Bradyrhizobium, Devosia, Rhodospirillium, ...*	---
Lip120_rev	3	LipH_f_3_20c	---	*Rhodopseudomonas, Bradyrhizobium, Devosia, Rhodospirillium, ...*	---
Lip129_rev	4	LipH_f_4_9c		Uncultured organism clone pLE03 hypothetical protein gene (800 bp)	---
Lip132_for	4	LipH_f_4_18c	*Azoarcus, Cupriavidus, Ralstonia, Burkoderia*	*Bradyrhizobium, Caulobacter, Rhodopseudomonas* (1.5 kbp), *Azoarcus, Cupriavidus, Burkoderia* (500bp)	---
Lip132_rev	4	LipH_f_4_18c	*Azoarcus, Cupriavidus, Ralstonia, Burkoderia*	*Bradyrhizobium, Caulobacter, Rhodopseudomonas* (1.5 kbp), *Azoarcus, Cupriavidus, Burkoderia* (500bp)	---
Lip137_for	4	LipH_f_4_1c	*Pseudomonas sp.*	*Burkholderia sp.* (1.2 kb), *Mycobacterium sp.* (1.3 kb)	---
Lip137_rev	4	LipH_f_4_1c	*Chromobacterium, Cupriavidus, Ralstonia, Pseudomonas*	*Burkholderia sp.* (1.2 kb), *Mycobacterium sp.* (1.3 kb)	---
Lip139_for	4	LipH_f_4_15c	---	---	---
Lip139_rev	4	LipH_f_4_15c	---	---	---
Lip142_for	4	LipH_f_4_29c	*Micromonospora sp.* (300 bp)	---	---
Lip142_rev	4	LipH_f_4_20c		---	---

PUBLICATIONS

STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

Received 28 September 2018
Accepted 25 February 2019

Edited by M. J. Romao, Universidade Nova de Lisboa, Portugal

Keywords: *Ignicoccus hospitalis*; α/β-hydrolases; metallo-β-lactamases.

Supporting information: this article has supporting information at journals.iucr.org/f

Igni18, a novel metallo-hydrolase from the hyperthermophilic archaeon *Ignicoccus hospitalis* KIN4/I: cloning, expression, purification and X-ray analysis

Stefanie Kobus,[a] Pablo Perez-Garcia,[b] Astrid Hoeppner,[a] Nicholas Holzscheck,[b] Filip Kovacic,[c] Wolfgang R. Streit,[b] Karl-Erich Jaeger,[c,d] Jennifer Chow[b] and Sander H. J. Smits[a,e]*

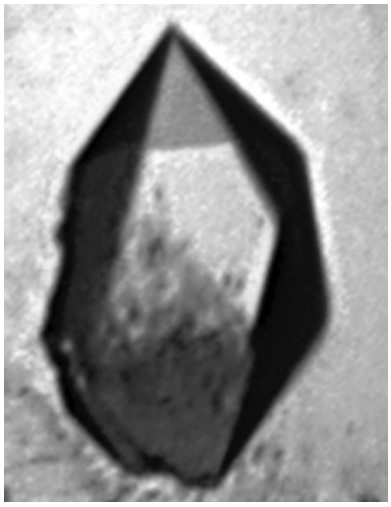
[a]Center for Structural Studies (CSS), Heinrich Heine University Düsseldorf, Universitätsstrasse 1, 40225 Düsseldorf, Germany, [b]Department of Microbiology and Biotechnology, University of Hamburg, Ohnhorststrasse 18, 22609 Hamburg, Germany, [c]Institute of Molecular Enzyme Technology (IMET), Heinrich Heine University Düsseldorf, 52426 Jülich, Germany, [d]Institute of Bio- and Geosciences IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, 52426 Jülich, Germany, and [e]Institute of Biochemistry, Heinrich Heine University Düsseldorf, Universitätsstrasse 1, 40225 Düsseldorf, Germany. *Correspondence e-mail: sander.smits@hhu.de

The hyperthermophilic crenarchaeon *Ignicoccus hospitalis* KIN4/I possesses at least 35 putative genes encoding enzymes that belong to the α/β-hydrolase superfamily. One of those genes, the metallo-hydrolase-encoding *igni18*, was cloned and heterologously expressed in *Pichia pastoris*. The enzyme produced was purified in its catalytically active form. The recombinant enzyme was successfully crystallized and the crystal diffracted to a resolution of 2.3 Å. The crystal belonged to space group *R*32, with unit-cell parameters $a = b = 67.42$, $c = 253.77$ Å, $\alpha = \beta = 90.0$, $\gamma = 120.0°$. It is suggested that it contains one monomer of Igni18 within the asymmetric unit.

1. Introduction

Ignicoccus hospitalis KIN4/I is an anaerobic, hyperthermophilic crenarchaeon with an optimal growth temperature of 90°C (363 K). It was isolated from a submarine hydrothermal system at the Kolbeinsey Ridge in the north of Iceland (Paper *et al.*, 2007). The archaeon belongs to the order Desulfurococcales and is an obligate chemolithoautotroph that grows by performing sulfur reduction using hydrogen as the electron donor. Among other uncommon metabolic pathways, *I. hospitalis* utilizes a rare autotrophic CO_2-fixation pathway starting from acetyl-coenzyme A (Jahn *et al.*, 2007). With 1.3 Mbp coding for 1444 proteins, it has one of the smallest genomes described for a free-living organism (Podar *et al.*, 2008). It is also the only organism known to date that is capable of acting as a host for *Nanoarchaeum equitans*, which is so far the only cultivated member of the Nanoarchaeota (Paper *et al.*, 2007; Huber *et al.*, 2000). This is the only natural archaeon–archaeon interaction that has been described *in vivo* (Wrede *et al.*, 2012), although it remains unclear whether it is a true mutualistic symbiosis or parasitism (Jahn *et al.*, 2008). Owing to the small genome size of *N. equitans* (0.5 Mbp), *I. hospitalis* provides the biological macromolecules that it cannot synthesize owing to a lack of essential biosynthesis pathways such as those for lipids, amino acids and nucleotides (Waters *et al.*, 2003) and even energetic precursors (Giannone *et al.*, 2011, 2015).

Living at above 90°C requires adaptation of all enzymes in order to carry out the metabolic reactions that are necessary



for existence. The amino-acid composition, especially a decrease in thermolabile residues such as asparagine and cysteine, hydrophobic interactions, aromatic interactions, ion pairs and increased salt-bridge networks, oligomerization and intersubunit interactions, packing and reduction of solvent-exposed surface area, flexibility of surface-exposed loops, and metal binding or substrate stabilization are some of the factors that have been attributed to confer protein stability at extreme temperatures (Unsworth *et al.*, 2007; Kovacic *et al.*, 2016).

Enzymes that belong to the α/β-hydrolase superfamily include metabolically important enzymes that carry out a large number of different hydrolysis and synthesis reactions. The metallo-β-lactamase (MBL) subfamily includes enzymes that hydrolyze thiol-ester, phosphodiester and sulfuric ester bonds, but also includes oxidoreductases. Many members of this subfamily are involved in mRNA maturation and DNA repair (Bebrone, 2007). The presence of MBL genes within the Eubacteria, Archaea and Eukaryota suggests a very ancient origin of this family (Garau *et al.*, 2005). The first crystal structure of an archaeal metallo-β-lactamase was published in 2010 (ST1585 from *Sulfolobus tokodaii*; Shimada *et al.*, 2010). MBL enzymes usually show a characteristic $\alpha\beta/\beta\alpha$ protein fold and are often dependent on Zn^{2+} ions (Meini *et al.*, 2015).

To date, only four crystal structures of native *I. hospitalis* proteins are available: a cell-adhesion structural protein (PDB entry 3j1r; Yu *et al.*, 2012), a membrane-associated octaheme cytochrome *c* (PDB entry 4q05; Parey *et al.*, 2016), the type IV pilus-like filament protein Iho670 (PDB entry 5kyh; Braun *et al.*, 2016) and a superoxide reductase (PDB entry 4bk8; Romão *et al.*, 2018). Here, we present the recombinant production, purification and crystallization of an MBL-like domain-containing protein from *I. hospitalis* KIN4/I.

2. Materials and methods

2.1. Cloning and overexpression

The *igni18* gene (NCBI accession No. IGNI_RS06455; start nucleotide No. 1115579, end nucleotide No. 1114878), which has been annotated as a metallo-hydrolase containing a metallo-β-lactamase fold (Podar *et al.*, 2008), was amplified with Phusion High-Fidelity DNA Polymerase (ThermoFisher Scientific, Carlsbad, California, USA) from *I. hospitalis* KIN4/I genomic DNA kindly donated by Dr Harald Huber (University of Regensburg, Germany). The EasySelect *Pichia* Expression Kit (Invitrogen, Carlsbad, California, USA) was employed for cloning and expression. The oligonucleotides designed for the amplification of *igni18* (Table 1) allowed the hydrolysis of the PCR product with EcoRI and NotI and cloning into the pPICZ-A expression plasmid (Invitrogen) using *Escherichia coli* DH5α. The start of the gene was modified to include a Kozak consensus sequence for yeast [5′-(G/A)NNATGG-3′; Romanos *et al.*, 1992] and the stop codon was omitted for fusion with a C-terminal Myc epitope and His_6 tag encoded by the vector sequence. These allow immunodetection and affinity-chromatographic purification of the produced fusion protein.

Table 1
Igni18 production information.

Source organism	*I. hospitalis* KIN4/I
DNA source	*I. hospitalis* KIN4/I
Forward primer† (5′–3′)	CCGAGAATTC**GACA**TG**G**CCACGGTTAAGCT GACCTAC
Reverse primer† (5′–3′)	AGCGGCCGCAAAATTCGAAGGTCACCGTCT CC
Cloning and expression vector	pPICZ-A
Cloning host	*E. coli* DH5α
Expression host	*P. pastoris* X-33
Igni18_Myc_His_6 amino-acid sequence	MATVKLTYFGHSAFHVEVDGVGIAIDPWIT NPLSKTTLEDYLKNFKTDLVVITHAHED HIGDALEIMRRTGAKFFSIHEIYVDLTQ KGFQGIGANIGGPAKLDDVAPGLGIALT PATHSSYDKGVPTGAIIFKDGKALVYHA GDTGLFAEMQFIGELYAPKVALLPIGGH YTMDIEQALLATKLLRPEVVVPMHYNTF PPIRADPNEFKQKVESAGLAKVRVMEPG ETVTFEFCGRQLGPEQKLISEEDLNSAV DHHHHHH

† Restriction sites are underlined and modifications are in bold.

The constructed pPICZ-A::*igni18* expression vector purified from *E. coli* DH5α was linearized with MssI restriction endonuclease (ThermoFisher Scientific) prior to electroporation of *Pichia pastoris* X-33 competent host cells (Invitrogen) according to the manual. Insertion into the chromosome occurs via homologous recombination at the 5′ AOX1 (alcohol oxidase 1) region. Colonies carrying the construct appeared on YPD–agar plates containing 1 *M* sorbitol and 100 µg ml^{-1} zeocin (Invitrogen) after 3–5 d of incubation at 303 K. The clones were tested for multiple insertions of the construct by the ability to grow on YPD–agar containing 1 mg ml^{-1} zeocin. A total of eight multi-insertion clones were tested for expression, and the production of the protein fused with a His_6 tag was verified by Western blotting using anti-His_6-tag-specific antibodies as described in the manual (Invitrogen). The clone yielding the highest Igni18 production was selected. It showed the methanol-utilization slow (Mut^S) phenotype, in which one of the copies of pPICZ-A::*igni18* replaced the original AOX1 locus, creating a mutant that can assimilate methanol only at slow rates. Fermentation was performed in 1.5 l buffered extra-YNB glycerol methanol (BYGM) auto-induction medium (Lee *et al.*, 2017) in a fermenter (Minifors, INFORS AG, Bottmingen, Switzerland) for 45 h at 303 K. Glycerol is used as the preferable carbon source for initial growth. After 24 h, the glycerol is consumed and methanol induces the expression of the gene of interest. The cells were harvested by centrifugation and the cell pellet was stored at 193 K until further use. Macromolecule-production information is summarized in Table 1.

2.2. Protein purification and enzyme-activity assays

For immobilized metal ion-affinity chromatography (IMAC) purification of Igni18_Myc_His_6, the cells were thawed on ice and suspended in 5 ml lysis buffer [10 mg ml^{-1} myristyl sulfobetaine (SB3-14), 1 m*M* phenylmethylsulfonyl fluoride (PMSF), 0.05 *M* NaH_2PO_4, 0.3 *M* NaCl pH 8.0] per gram. Cell disruption and partial purification were achieved by

Table 2
Crystallization conditions.

Method	Sitting-drop vapor diffusion
Plate type	MRC3, Swissci
Temperature (K)	293
Protein concentration (mg ml^{-1})	20
Buffer composition of protein solution	0.1 *M* potassium phosphate pH 7
Composition of reservoir solution	0.3 *M* magnesium nitrate hexahydrate, 0.1 *M* Tris pH 8, 22%(*w*/*v*) PEG 8000
Volume and ratio of drop	200 nl, 1:1
Volume of reservoir (µl)	40

incubating the cells at 343 K for 1 h in the presence of the zwitterionic detergent SB3-14 (Zanna & Haeuw, 2007). Cell debris was removed from the crude cell extract by centrifugation at 15 000 rev min^{-1} for 30 min at 277 K (Sorvall RC6+ centrifuge, SS-34 rotor; Thermo Scientific, Braunschweig, Germany). The clear lysate was loaded onto a Protino Ni-TED 2000 Packed Column (Macherey-Nagel, Düren, Germany) and Igni18 was purified according to the manufacturer's instructions. The buffer of the elution fractions was exchanged to 0.1 *M* potassium phosphate buffer pH 7.0 using an ultrafiltration unit with a pore size of 10 kDa (Vivaspin 20, Sartorius AG, Göttingen, Germany). The proteins were analyzed by polyacrylamide gel electrophoresis under denaturing conditions (SDS–PAGE) on 12%(*w*/*v*) gels stained with Coomassie Brilliant Blue G-250 (Laemmli, 1970) and by Western blotting (Fig. 1). Activity was assayed on *para*-nitrophenyl (*p*NP) esters with fatty-acid chain lengths ranging from two to 18 C atoms (Jaeger & Kovacic, 2014).

2.3. Crystallization and preliminary X-ray analysis of Igni18

Crystallization trials were performed using the sitting-drop vapor-diffusion method at 293 and 285.15 K. To find an initial crystallization condition, various commercial kits from Qiagen (Hilden, Germany) and Molecular Dimensions (Suffolk, England) were used in MRC3 Swissci plates.

0.1 µl of homogenous recombinant Igni18 (at either 15 or 20 mg ml^{-1} in 0.1 *M* potassium phosphate buffer pH 7) was mixed with 0.1 µl reservoir solution and equilibrated against 40 µl reservoir solution using a pipetting robot (NT8, Formulatrix, Bedford, Massachusetts, USA).

Crystals were grown at room temperature and reached final dimensions of around 130 × 90 × 80 µm after several months (Fig. 2*a*). The crystallization condition consisted of 0.3 *M* magnesium nitrate hexahydrate, 0.1 *M* Tris pH 8, 22%(*w*/*v*) PEG 8000. To verify that the crystals contained protein, a fluorescence image was obtained using the intrinsic fluorescence of the tryptophan residue present in the Igni18 protein (Fig. 2*b*). The crystal was cryoprotected by overlaying the drop with 2 µl mineral oil and subsequently flash-cooled in liquid nitrogen. Crystallization information is summarized in Table 2.

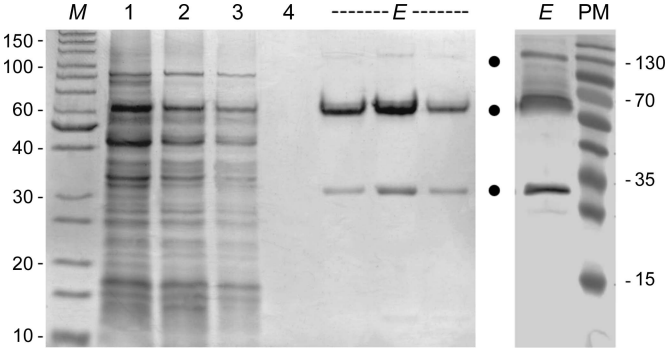
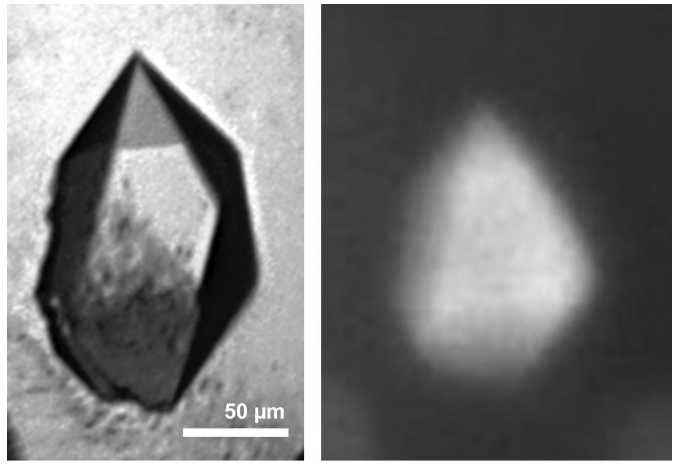


Figure 1
Ni-TED affinity purification of Igni18_Myc_His$_6$ (SDS–PAGE, left; Western blot, right). The samples were mixed with a loading dye containing 0.1 *M* dithiothreitol and were partially denatured at 368 K (95°C) for 10 min. Lane *M*, protein molecular-weight marker (PageRuler Unstained Protein Ladder, Thermo Scientific, Braunschweig, Germany; labeled in kDa); lane 1, lysate; lane 2, flowthrough; lane 3, initial column wash; lane 4, final column wash; lane *E*, elution fraction; lane PM, prestained protein molecular-weight marker (PageRuler Prestained Protein Ladder, Thermo Scientific, Braunschweig, Germany; labeled in kDa). Bands corresponding to monomeric Igni18 and oligomers are marked with dots and were confirmed by Western blot analysis using His$_6$-specific antibodies.

2.4. Data collection and processing

A data set was collected from a single Igni18 crystal on beamline ID30A-3 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France at 100 K. Initially, four frames were taken for characterization with 0.5° rotation at angles of 0°, 45°, 90° and 135°. The images were processed and a strategy was calculated using *BEST* (Bourenkov & Popov, 2010). A data set was then collected based on this strategy, resulting in the use of 0.05° rotation per frame with a total of 2500 frames (a total of 125° rotation). This data set was initially processed using the automated processing pipeline at the ID30A-3 beamline and was subsequently reprocessed using *XDS* (Kabsch, 2010), *XSCALE* and *POINTLESS* (Evans, 2006) to determine the space group. Data-collection and processing statistics are summarized in Table 3. To check for anomalous signal, the data were separately processed such that the Friedel pairs remained unmerged.



(*a*) (*b*)

Figure 2
Crystal of Igni18. Visible image (*a*) and UV image (*b*) of an Igni18 crystal obtained in 0.3 *M* magnesium nitrate hexahydrate, 0.1 *M* Tris pH 8, 22%(*w*/*v*) PEG 8000.

Table 3
Data collection and processing.

Values in parentheses are for the highest resolution shell.

X-ray source	ID30A-3, ESRF, Grenoble
Detector	EIGER 4M
Wavelength (Å)	0.9677
Temperature (K)	100
Crystal-to-detector distance (mm)	118.16
Rotation range per image (°)	0.05
Total rotation range (°)	125
Exposure time (s)	0.002
Space group	*R*32
a, *b*, *c* (Å)	67.42, 67.42, 253.77
α, *β*, *γ* (°)	90.0, 90.0, 120.0
Resolution range (Å)	31.32–2.30 (2.382–2.300)
Total reflections	70412 (7119)
Unique reflections	10293 (1008)
R_{merge}	0.1004 (0.5445)
R_{meas}	0.1084 (0.5872)
$R_{p.i.m.}$	0.03987 (0.2152)
$CC_{1/2}$	0.998 (0.885)
Wilson *B* factor (Å^2)	32.46
$\langle I/\sigma(I)\rangle$	12.60 (3.37)
Completeness (%)	99.78 (100.00)
Multiplicity	6.8 (7.1)
Matthews coefficient (Å^3 Da^{-1})	2.17
Solvent content (%)	43.4

3. Results and discussion

A *BLASTp* search (Altschul *et al.*, 2005) against the GenBank nonredundant database indicated that enzymes similar to Igni18 can only be found within the genus *Ignicoccus* with 78% identity. Furthermore, less similar enzymes with an identity of 52% or lower mostly belong to other members of the Crenarchaeota. The physiological function of this lipase and putative metallo-hydrolase in the respective hyperthermophilic organisms needs to be elucidated. To date, only about 61% of the protein-coding genes within *I. hospitalis* KIN4/I have bioinformatically predicted functions, implying that many pathways within this archaeon as well as its symbiont *N. equitans* are still unknown.

Initial attempts to produce the metallo-hydrolase Igni18 in different *E. coli* strains failed (data not shown); therefore, we used the yeast *P. pastoris* carrying multiple chromosomally integrated copies of the *igni18* gene as an expression system. It seems that some of the problems regarding translation and post-translational modifications of the archaeal protein can be circumvented by using the eukaryotic *P. pastoris* compared with *E. coli*. After purification by affinity chromatography on Ni-TED agarose, the purity of Igni18_Myc_His_6 was estimated to be greater than 90% and the yield was approximately 8 mg per litre of expression culture. The purified Igni18_Myc_His_6 monomer has an apparent molecular weight of 30 kDa. which is in good agreement with the theoretical value of 28 614 Da (Fig. 1). The native protein seems to occur in an oligomeric state and this complex is highly stable, as expected from the nature of its native host; it can be denatured only partially with a reducing loading dye and heat incubation for 10 min at 368 K (95°C). Using a Western blot immunodetection method, the three bands could be identified as His_6-tagged Igni18 (Fig. 1). Under semi-native running conditions, *i.e.* SDS–PAGE with nonreducing loading dye and without prior heat denaturation, only one single band could be observed on the gel at slightly below 130 kDa (data not shown). This suggests that native, soluble Igni18 predominantly exists in a multimeric form.

The enzyme activity of Igni18 was confirmed on *p*NP esters. Further assays are required with lipase-specific and MBL-specific substrates and need to be performed to reveal the precise biochemical function of Igni18. The optimal enzymatic activity of Igni18 was determined as 363 K (90°C), which corresponds to the optimal growth temperature of the host organism (Pérez-García *et al.*, unpublished work).

The purified Igni18 was used to screen for suitable crystallization conditions using different commercial screening kits from Qiagen (Hilden, Germany) and Molecular Dimensions (Suffolk, England) and an NT8 pipetting robot (Formulatrix). Only a few Igni18 crystals appeared after several months in a condition consisting of 0.3 *M* magnesium nitrate hexahydrate, 0.1 *M* Tris pH 8, 22%(*w*/*v*) PEG 8000. Although the crystallization condition was further optimized by grid screening, only the crystal obtained in the initial screen displayed diffraction quality suitable for the collection of a data set.

From this crystal, which diffracted to beyond 2.0 Å resolution (Fig. 3), a complete data set was collected on beamline ID30A-3 at the ESRF, Grenoble, France to a resolution of 2.3 Å. Here, we cut the data owing to incompleteness at higher resolutions. Preliminary X-ray diffraction analysis showed that the Igni18 crystal belonged to the hexagonal space group *R*32, with unit-cell parameters $a = b = 67.42$, $c = 253.77$ Å, $\alpha = \beta = 90.0$, $\gamma = 120.0°$. Calculation of the unit-cell volume indicated the presence of one monomer in the asymmetric unit with a

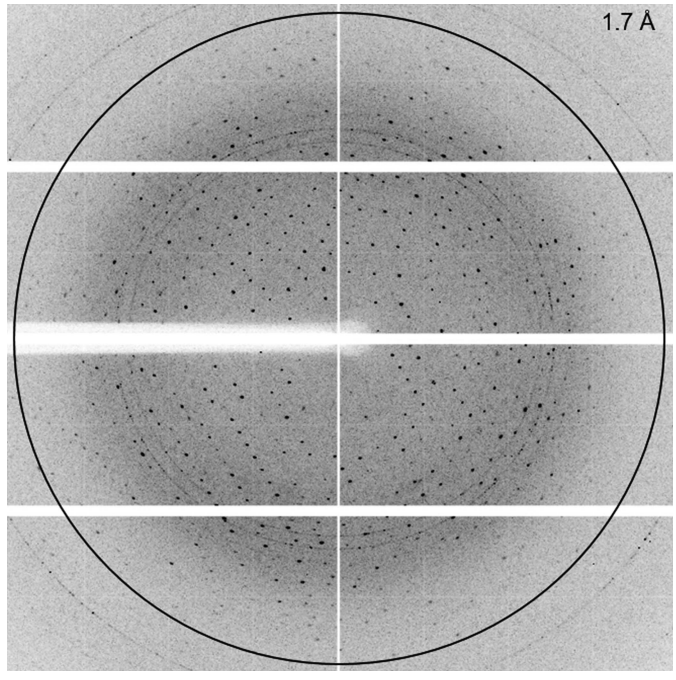


Figure 3
Diffraction image of an Igni18 crystal with a resolution of 1.7 Å at the edge of the detector (indicated by the black circle).

calculated V_M of 2.17 Å^3 Da^{-1} and a solvent content of 43.4% (Matthews, 1968).

Interestingly, the anomalous signal after processing using *XDS* and *XSCALE* revealed the presence of an anomalous scatterer in the protein crystal, which was likely to result from endogenously bound metal within the Igni18 protein. Since Igni18 is a metallo-hydrolase, this suggests that this anomalous scatterer was bound to the protein directly during expression. Using the anomalous scattering analysis server (http://skuld.bmsc.washington.edu/scatter/), we analyzed which scatterer might be present. At the used wavelength of 0.9677 Å, we assume that manganese, iron, cobalt, nickel, copper and zinc might be possible candidates. They display scattering coefficients f' or f'' which range from 1.4 e for manganese to 3.9 e for iron. Since we did not add any of these ions to the buffers or the protein, we cannot precisely tell which one was bound by the Igni18 protein. By sequence homology, this protein belongs to the archaeal metallo-β-lactamase family, suggesting that it would be functional with zinc. However, we used a nickel column during purification, so this also might be the bound scatterer. Initial substructure analysis using *AutoSol* as part of the *PHENIX* suite (Adams *et al.*, 2010) indicates that two atoms are bound within the asymmetric unit, suggesting, in combination with the Matthews coefficient, that one monomer has two anomalous scatterers bound. We used *Auto-Rickshaw* (Panjikar *et al.*, 2009) to obtain the initial phases using the MR-SAD option, supplying only the Igni18 protein sequence as model input, which resulted in an initial model which is currently being refined. Although a monomer is present in the asymmetric unit, a trimeric Igni18 protein is likely to be built by the threefold axes given by the *R*32 symmetry. This suggests that despite the presence of only one monomer in the asymmetric unit the Igni18 protein crystallized as an oligomeric protein. The electron density is of good quality and allowed the entire sequence of Igni18 to be fitted unambiguously (see Supplementary Fig. S1).

Acknowledgements

X-ray diffraction measurements were performed on beamline ID30A-3 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. We are grateful to Guillaume Gotthard at the ESRF for providing assistance in using beamline ID30A-3. We thank Dr Harald Huber, University of Regensburg, Germany for providing us with genomic DNA of *I. hospitalis*. This project was partially funded by the Horizon2020 project INMARE (Grant agreement ID: 634486).

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Cryst.* D**66**, 213–221.
Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schäffer, A. A. & Yu, Y.-K. (2005). *FEBS J.* **272**, 5101–5109.
Bebrone, C. (2007). *Biochem. Pharmacol.* **74**, 1686–1701.
Bourenkov, G. P. & Popov, A. N. (2010). *Acta Cryst.* D**66**, 409–419.
Braun, T., Vos, M. R., Kalisman, N., Sherman, N. E., Rachel, R., Wirth, R., Schoeder, G. F. & Egelman, E. H. (2016). *Proc. Natl Acad. Sci. USA*, **113**, 10352–10357.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Garau, G., Di Guilmi, A. M. & Hall, B. G. (2005). *Antimicrob. Agents Chemother.* **49**, 2778–2784.
Giannone, R. J., Huber, H., Karpinets, T., Heimerl, T., Küper, U., Rachel, R., Keller, M., Hettich, R. L. & Podar, M. (2011). *PLoS One*, **6**, e22942.
Giannone, R. J., Wurch, L. L., Heimerl, T., Martin, S., Yang, Z., Huber, H., Rachel, R., Hettich, R. L. & Podar, M. (2015). *ISME J.* **9**, 101–114.
Huber, H., Burggraf, S., Mayer, T., Wyschkony, I., Rachel, R. & Stetter, K. O. (2000). *Int. J. Syst. Evol. Microbiol.* **50**, 2093–2100.
Jaeger, K. E. & Kovacic, F. (2014). *Methods Mol. Biol.* **1149**, 111–134.
Jahn, U., Gallenberger, M., Paper, W., Junglas, B., Eisenreich, W., Stetter, K. O., Rachel, R. & Huber, H. (2008). *J. Bacteriol.* **190**, 1743–1750.
Jahn, U., Huber, H., Eisenreich, W., Hügler, M. & Fuchs, G. (2007). *J. Bacteriol.* **189**, 4108–4119.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kovacic, F., Mandrysch, A., Poojari, C., Strodel, B. & Jaeger, K. E. (2016). *Protein Eng. Des. Sel.* **29**, 65–76.
Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
Lee, J. Y., Chen, H., Liu, A., Alba, B. M. & Lim, A. C. (2017). *Protein Expr. Purif.* **137**, 7–12.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Meini, M.-R., Llarrull, L. I. & Vila, A. J. (2015). *FEBS Lett.* **589**, 3419–3432.
Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2009). *Acta Cryst.* D**65**, 1089–1097.
Paper, W., Jahn, U., Hohn, M. J., Kronner, M., Näther, D. J., Burghardt, T., Rachel, R., Stetter, K. O. & Huber, H. (2007). *Int. J. Syst. Evol. Microbiol.* **57**, 803–808.
Parey, K., Fielding, A. J., Sörgel, M., Rachel, R., Huber, H., Ziegler, C. & Rajendran, C. (2016). *FEBS J.* **283**, 3807–3820.
Podar, M., Anderson, I., Makarova, K. S., Elkins, J. G., Ivanova, N., Wall, M. A., Lykidis, A., Mavromatis, K., Sun, H., Hudson, M. E., Chen, W., Deciu, C., Hutchison, D., Eads, J. R., Anderson, A., Fernandes, F., Szeto, E., Lapidus, A., Kyrpides, N. C., Saier, M. H. Jr, Richardson, P. M., Rachel, R., Huber, H., Eisen, J. A., Koonin, E. V., Keller, M. & Stetter, K. O. (2008). *Genome Biol.* **9**, R158.
Romanos, M. A., Scorer, C. A. & Clare, J. J. (1992). *Yeast*, **8**, 423–488.
Romão, C. V., Matias, P. M., Sousa, C. M., Pinho, F. G., Pinto, A. F., Teixeira, M. & Bandeiras, T. M. (2018). *Biochemistry*, **57**, 5271–5281.
Shimada, A., Ishikawa, H., Nakagawa, N., Kuramitsu, S. & Masui, R. (2010). *Proteins*, **78**, 2399–2402.
Unsworth, L. D., van der Oost, J. & Koutsopoulos, S. (2007). *FEBS J.* **274**, 4044–4056.
Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., Beeson, K. Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E., Ni, J., Podar, M., Richardson, T., Sutton, G. G., Simon, M., Söll, D., Stetter, K. O., Short, J. M. & Noordewier, M. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 12984–12988.
Wrede, C., Dreier, A., Kokoschka, S. & Hoppert, M. (2012). *Archaea*, **2012**, 596846.
Yu, X., Goforth, C., Meyer, C., Rachel, R., Wirth, R., Schröder, G. F. & Egelman, E. H. (2012). *J. Mol. Biol.* **422**, 274–281.
Zanna, L. & Haeuw, J. F. (2007). *J. Chromatogr. B*, **846**, 368–373.

ACS chemical biology

Articles

Cite This: ACS Chem. Biol. 2018, 13, 225−234

Determinants and Prediction of Esterase Substrate Promiscuity Patterns

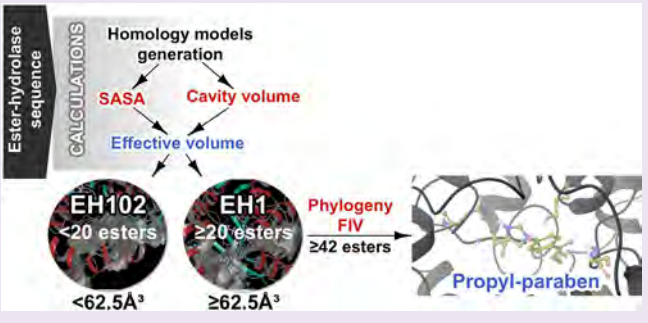
Mónica Martínez-Martínez,[†,ζ] Cristina Coscolín,[†,ζ] Gerard Santiago,[‡,ζ] Jennifer Chow,[§] Peter J. Stogios,[∥] Rafael Bargiela,[†,γ] Christoph Gertler,[⊥,δ] José Navarro-Fernández,[†] Alexander Bollinger,[#] Stephan Thies,[#] Celia Méndez-García,[▽,ϵ] Ana Popovic,[∥] Greg Brown,[∥] Tatyana N. Chernikova,[⊥] Antonio García-Moyano,[○] Gro E. K. Bjerga,[○] Pablo Pérez-García,[§] Tran Hai,[⊥] Mercedes V. Del Pozo,[†] Runar Stokke,[◆] Ida H. Steen,[◆] Hong Cui,[∥] Xiaohui Xu,[∥] Boguslaw P. Nocek,[¶] María Alcaide,[†] Marco Distaso,[⊥] Victoria Mesa,[▽] Ana I. Peláez,[▽] Jesús Sánchez,[▽] Patrick C. F. Buchholz,[%] Jürgen Pleiss,[%] Antonio Fernández-Guerra,[$,●,■] Frank O. Glöckner,[$,●] Olga V. Golyshina,[⊥] Michail M. Yakimov,[⌀,★] Alexei Savchenko,[∥] Karl-Erich Jaeger,[#,α] Alexander F. Yakunin,[∥,ψ] Wolfgang R. Streit,[§,ψ] Peter N. Golyshin,[⊥,ψ] Víctor Guallar,*[,‡,β,ψ] Manuel Ferrer,*[,†,ψ] and The INMARE Consortium

[†]Institute of Catalysis, Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain
[‡]Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain
[§]Biozentrum Klein Flottbek, Mikrobiologie & Biotechnologie, Universität Hamburg, 22609 Hamburg, Germany
[∥]Department of Chemical Engineering and Applied Chemistry, University of Toronto, M5S 3E5 Toronto, Ontario, Canada
[⊥]School of Biological Sciences, Bangor University, LL57 2UW Bangor, United Kingdom
[#]Institut für Molekulare Enzymtechnologie, Heinrich-Heine-Universität Düsseldorf, 52425 Jülich, Germany
[▽]Department of Functional Biology-IUBA, Universidad de Oviedo, 33006 Oviedo, Spain
[○]Uni Research AS, Center for Applied Biotechnology, 5006 Bergen, Norway
[◆]Department of Biology and KG Jebsen Centre for Deep Sea Research, University of Bergen, 5020 Bergen, Norway
[¶]Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, 60439 Illinois, United States
[%]Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, 70569 Stuttgart, Germany
[$]Jacobs University Bremen gGmbH, Bremen, Germany
[●]Max Planck Institute for Marine Microbiology, 28359 Bremen, Germany
[■]University of Oxford, Oxford e-Research Centre, Oxford, United Kingdom
[⌀]Institute for Coastal Marine Environment, Consiglio Nazionale delle Ricerche, 98122 Messina, Italy
[★]Immanuel Kant Baltic Federal University, 236041 Kaliningrad, Russia
[α]Institute for Bio- and Geosciences IBG-1: Biotechnology, Forschunsgzentrum Jülich GmbH, 52425 Jülich, Germany
[β]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

Supporting Information

ABSTRACT: Esterases receive special attention because of their wide distribution in biological systems and environments and their importance for physiology and chemical synthesis. The prediction of esterases' substrate promiscuity level from sequence data and the molecular reasons why certain such enzymes are more promiscuous than others remain to be elucidated. This limits the surveillance of the sequence space for esterases potentially leading to new versatile biocatalysts and new insights into their role in cellular function. Here, we performed an extensive analysis of the substrate spectra of

continued...



Received: November 21, 2017
Accepted: November 28, 2017
Published: November 28, 2017

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

145 phylogenetically and environmentally diverse microbial esterases, when tested with 96 diverse esters. We determined the primary factors shaping their substrate range by analyzing substrate range patterns in combination with structural analysis and protein−ligand simulations. We found a structural parameter that helps rank (classify) the promiscuity level of esterases from sequence data at 94% accuracy. This parameter, the active site effective volume, exemplifies the topology of the catalytic environment by measuring the active site cavity volume corrected by the relative solvent accessible surface area (SASA) of the catalytic triad. Sequences encoding esterases with active site effective volumes (cavity volume/SASA) above a threshold show greater substrate spectra, which can be further extended in combination with phylogenetic data. This measure provides also a valuable tool for interrogating substrates capable of being converted. This measure, found to be transferred to phosphatases of the haloalkanoic acid dehalogenase superfamily and possibly other enzymatic systems, represents a powerful tool for low-cost bioprospecting for esterases with broad substrate ranges, in large scale sequence data sets.

Enzymes with outstanding properties in biological systems and the conditions favoring their positive selection are difficult to predict. One of these properties is substrate promiscuity, which typically refers to a broad substrate spectrum and acceptance of larger substrates. This phenomenon is important from environmental,[1] evolutionary,[2−5] structural,[6−8] and biotechnological[9,10] points of view. The relevance of substrate promiscuity is indisputable as the operating basis for biological processes and cell function. As an example, the evolutionary progress of enzymes from lower to higher substrate specificity allows the recruitment of alternate pathways for carbon cycling and innovations across metabolic subsystems and the tree of life by maximizing the growth rate and growth efficiency.[11] Promiscuous enzymes are energetically more favorable than specialized enzymes,[4] and therefore, the cell does not require many different enzymes to take up substrates, favoring genome minimization and streamlining.[12] In addition, the acquisition of new specificities without compromising primary or ancestral ones is a major driver of microbial adaptation to extreme habitats.[13] From a more practical standpoint, along with requirements of a technical nature such as selectivity, scalability and robustness, a narrow substrate spectrum is one of the most frequent problems for industrial enzyme applications.[14] A consensus exists that "the more substrates an enzyme converts the better," opening application ranges with consequent reduction of the production cost of multiple enzymes.[10,14,15]

Enzymes with wide substrate ranges occur naturally, as systematically investigated for halo-alkane dehalogenases,[16] phosphatases,[1] beta-lactamases,[2,17] and hydroxyl-nitrile lyases.[5] Some enzymes are more promiscuous than others simply due to their fold or degree of plasticity or the presence of structural elements or mutations occurring under selection in the proximity of the active-site cavity and access tunnels favoring promiscuity. However, the general explanation, if any, by which an enzyme binds and converts multiple substrates is unknown, although molecular insights have been reported for single enzymes.[18] A tool that can clearly distinguish promiscuous versus non-promiscuous enzymes and suggest substrates potentially being converted or not by them might therefore be valuable in applying low-cost sequencing in discovery platforms in any biological context.

In an ideal scenario, functional characterization of enzymes with genomics[19] and metagenomics[10,20] techniques using a large library of substrates would guide the analysis of sequence-to-promiscuity relationships and explore the mechanistic basis of promiscuity. In addition, such studies may help identify a new generation of highly promiscuous microbial biocatalysts. However, extensive bioprospecting and biochemical studies are rare,[10] despite the growing number of sequences available through low-cost sequencing efforts[21] and the growing number of enzymes that are typically characterized with limited substrate sets.[14] To address this knowledge gap, we functionally assessed the substrate specificity of a set of 145 phylogenetically, environmentally, and structurally diverse microbial esterases (herein referred to as "EH," which means Ester Hydrolase) against a customized library of 96 different substrates to find predictive markers of substrate promiscuity rather than discrete determinants of substrate specificity that may differ from protein to protein. EHs were selected for an analysis of substrate promiscuity because they typically have specific definitions of molecular function, can be easily screened in genomes and metagenomes compared with many other classes of proteins, are among the most important groups of biocatalysts for chemical synthesis, and are widely distributed in nature, with at least one EH per genome.[14]

Our work adds important insights and empirical, structural, and computational data to facilitate the elucidation of the molecular basis of substrate promiscuity in EHs, which was further extended to phosphatases from the haloalkanoic acid dehalogenase (HAD) superfamily. This was achieved by deciphering what we consider a predictive structural marker of substrate promiscuity and by establishing the reasons why certain such enzymes are more promiscuous than others and can convert substrates that others cannot. This study does not pretend to generate a quantitative measure to predict the number of compounds that an enzyme will hydrolyze but a tool and a parameter that will help in ranking (classifying) promiscuity level. Following on from that, we propose in this work the first molecular classification method of this kind derived from first principle molecular simulations and with clear physical/structural interpretation. This work also provides an example of the utility of this parameter to screen the sequence space for highly promiscuous EHs that may compete with best commercial EH preparations. We also provide first preliminary evidence of a number of underexplored microbial phylogenetic lineages containing EHs with a prominent substrate range.

RESULTS AND DISCUSSION

The Substrate Range of 145 Diverse EHs. A total of 145 EHs were investigated. Extensive details of the sources and screen methods are provided in the Supporting Information Methods and Table S1. In an environmental context, the source of enzymes was highly diverse because they were isolated from bacteria from 28 geographically distinct sites (125 EHs in total) and from six marine bacterial genomes (20 EHs; Supporting Information Figure S1). A phylogenetic analysis also indicated that sequences belong to bacteria distributed across the entire phylogenetic tree (Supporting Information Results and Figure S2).

The 145 putative proteins exhibited maximum amino acid sequence identities (Supporting Information Table S1) ranging from 29.1 to 99.9% to uncharacterized homologous proteins in public databases, with an average value (reported as %, with the

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

interquartile range (IQR) in parentheses) of 74.3% (40.3%). The pairwise amino acid sequence identity for all EHs ranged from 0.2 to 99.7% (Supporting Information Table S2), with an average value of 13.7% (7.6%). BLAST searches were performed for all query sequences by running NCBI BLASTP against the current version of the Lipase Engineering Database[22] using an E-value threshold of 10^{-10} and were successful for all but nine candidates. A total of 120 EH sequences were unambiguously assigned to some of the 14 existing families (F) of the Arpigny and Jaeger classification, which are defined based on amino acid sequence similarity and the presence of specific sequence motifs.[14,23] These EHs included sequences with a typical α/β hydrolase fold and conserved G-X-S-X-G (FI, 20; FIV, 36; FV, 33; FVI, 5; and FVII, 6) or G-X-S-(L) (FII, 9) motifs and sequences with a serine beta-lactamase-like modular (non-α/β hydrolase fold) architecture and a conserved S-X-X-K motif (FVIII, 11). An additional set of nine sequences were assigned to the *meta*-cleavage product (MCP) hydrolase family[24] and six to the so-called carbohydrate esterase family,[25] both with typical α/β hydrolase folds. Finally, one was a cyclase-like protein from the amido-hydrolase superfamily.[26] Sequences-to-family assignments are summarized in the Supporting Information Table S1. Taken together, the primary sequence analysis suggests that the diversity of polypeptides is not dominated by a particular type of protein or highly similar protein clusters but consists of diverse nonredundant sequences assigned to multiple folds and subfamilies, which are distantly related to known homologues in many cases.

The substrate profiles of all EHs were examined using a set of 96 chemically and structurally distinct esters (Supporting Information Table S3). We are aware that the number of compounds hydrolyzed may be an ambiguous indicator of promiscuity, because the size and composition of the library may influence the results. For this reason, the composition of the library was not random but based on including esters with variation in size of acyl and alcohol groups and with growing residues (aromatic, aliphatic, branched, and unbranched) at both sides, leading to more challenging substrates because a larger group adjacent to the ester bond increases the difficulty of conversion. Halogenated, chiral, and sugar esters, lactones, and an alkyl diester were also included. Esters with nitro substituents were not included. We used the partitioning coefficient (log P value) to indicate the chemical variability of the esters because this parameter reflects electronic and steric effects and hydrophobic and hydrophilic characteristics. Log P was determined with the software ACD/ChemSketch 2015.2.5. Log P values (Supporting Information, Table S3) ranged from −1.07 (for methyl glycolate) to 23.71 (for triolein), with an average value (IQR in parentheses) of 3.13 (2.86), which indicates that the ester library used in this study had broad chemical and structural variability. Nevertheless, adding new substrates could surely help (and even change) the ranking of the EHs herein analyzed. The dynamic range of the assay may also influence the results. For this reason, to detect enzyme−substrate pairs for a given EH, the ester library was screened with each of the 145 EHs in a kinetic pH indicator assay in 384-well plates,[24,27,28] which unambiguously allow quantifying specific activities at pH 8.0 and 30 °C, using a substrate concentration above 0.5 mM (see Supporting Information, Results). Two commercial lipases, CalA and CalB from *Pseudozyma aphidis* (formerly *Candida antarctica*), were included in the assays for comparison. Using this data set, we linked the biocatalytic data to the sequence information for the respective enzyme. In this study, sequence information meant any sequence that encoded an EH of interest. Biocatalytic data meant experimental data on substrate conversion (i.e., units g^{-1} or U g^{-1}) followed for 24 h.

We determined the probability of finding an EH with a broad substrate profile by plotting the number of esters that were hydrolyzed by all preparations. Figure 1 shows that the number

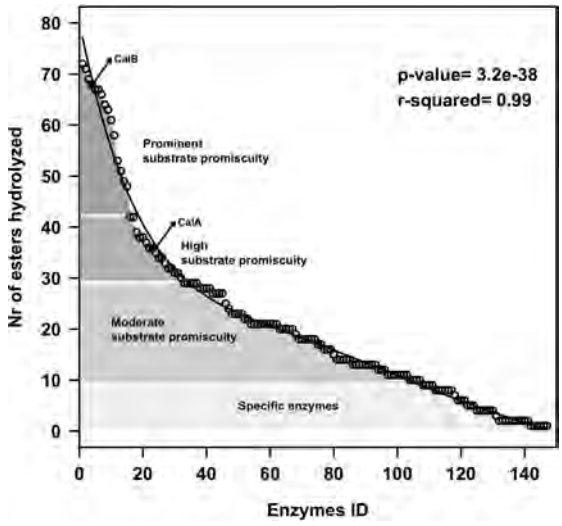


Figure 1. Number of ester substrates hydrolyzed by each of the 145 EHs investigated in this study. The commercial preparations CalA and CalB (marked with filled square) are also included. This figure is created from data in the Supporting Information Table S1. The activity protocol established and used to identify the esters hydrolyzed by each EH was based on a 550 nm follow-up pH indicator assay described in the Supporting Information Methods. The list of the 96 structurally different esters tested is shown in Figure 2. Full details of the activity protocol are provided in the Supporting Information Methods. The trend line shows a not-single exponential fit of the experimental data. The fit was obtained using R script and the "lm" function, to extract a polynomial regression of degree 6 with the following line "model ← lm(MM[,1] ~ poly(MM[,2],6,raw = TRUE))", where MM[,1] corresponds to the number of esters hydrolyzed, and MM[,2] the position in the *x* axis (from 1 to 147).

of esters hydrolyzed by all 147 EHs (including CalA/B) fits to an exponential distribution (r^2 = 0.99; p value $3.2e^{-38}$; Pearson's correlation coefficient) with a median of 18 substrates per enzyme, nine hits at the 25th percentile, and 29 hits at the 75th percentile. On the basis of this distribution and a previously established criterion,[1] we considered an enzyme specific if it used nine esters or fewer (27% of the total), as showing moderate substrate promiscuity if it used between 10 and 29 esters (51% of the total), and as showing high-to-prominent promiscuity if it used 30 or more esters (22% of the total). This criterion indicated a percentage of EHs with a prominent substrate range similar to that found for HAD phosphatases (24%).[1]

Phylogeny Is a Predictive Marker of Substrate Promiscuity. Hierarchical clustering was performed to evaluate the differences in substrate range patterns (Figure 2). For the sake of simplicity, clustering was performed for those EHs that hydrolyzed 10 or more esters (i.e., 107 total EHs). We first observed a large percentage of enzymes with presumptive broad active site environments that accommodated large aromatic and sterically hindered esters such as benzyl (*R*)-(+)-2-hydroxy-3-phenylpropionate (49% of the total), benzoic acid-4-formyl-phenylmethyl ester (27%), 2,4-dichlorophenyl 2,4-dichlorobenzoate (~8%), 2,4-dichlorophenyl 2,4-dichlorobenzoate (~5%), and diethyl-2,6-dimethyl 4-phenyl-1,4-dihydro

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

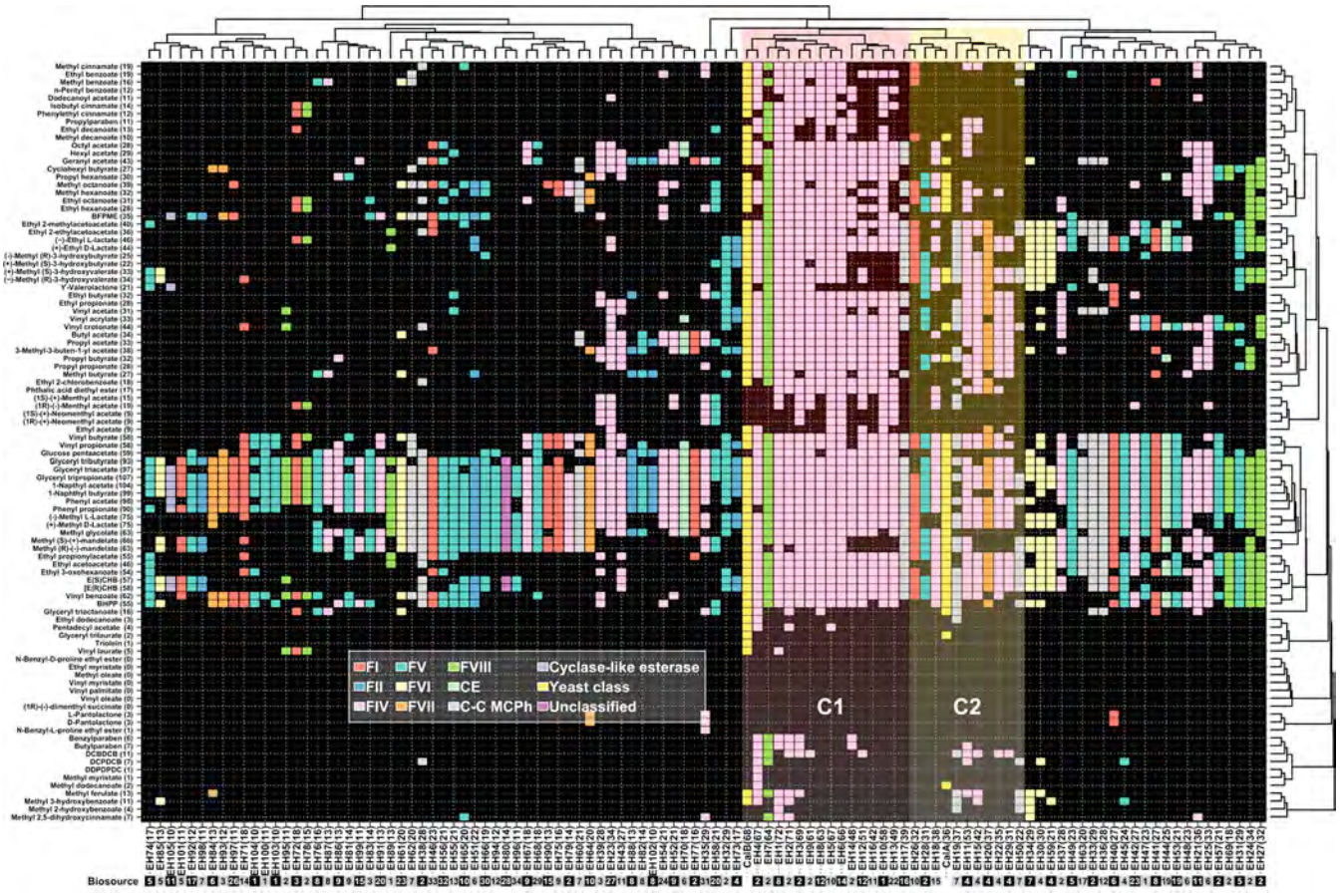


Figure 2. Hierarchical clustering of the substrate ranges of the EHs. Only EHs that hydrolyzed 10 or more esters were considered (107 in total, including CalA/B). This figure is created from data in the Supporting Information Table S3. The specific activities of the EHs for each of the 96 esters were determined as described in Figure 1. The list of the 96 esters tested and the frequency of each ester considered as a hit (in brackets) are shown on the left side. The ID code representing each EH is given at the bottom. Each hydrolase is named based on the code "EH," which means Ester Hydrolase, followed by an arbitrary number from 1 to 145 for the most to least promiscuous enzyme. The number in brackets indicates the number of esters hydrolyzed by each enzyme. The biosource of each EH is indicated at the bottom with a number in white or black squares that follows the nomenclature in the Supporting Information Figure S1. The figure was created with the R language console using a binomial table with information about the activity/inactivity (1/0) of the analyzed enzymes against the 96 substrates as a starting point. For the central graphic, which shows the data in Supporting Information Table S3, we used the drawing tools provided by the basic core packages of R. The hierarchical clusters of the enzymes (shown at the top) and substrates (shown on the right side) were generated by calculating a distance matrix using a "binomial" method and the hclust function to generate the tree. Using the functions as .phylo and plot.phylo from the ape package, the clusters were added to the top and right of the figure. A combination of the Set1 palette from the R package RColorBrewer and colors from the basic palette from R were used as the color palette for sequences assigned to each family (F; see inset), including FI to FVII, carbohydrate esterase (CE), and carbon−carbon *meta*-cleavage product hydrolase (C−C MCP) families, all with a typical α/β hydrolase fold, FVIII serine beta-lactamase with non α/β hydrolase fold, and cyclase-like protein from the amido-hydrolase superfamily. Sequences that were not unambiguously ascribed to existing families were referred to as "Unclassified," and those of yeast origin were assigned to "yeast class." The two "clusters" C1 and C2 that contained the most substrate-promiscuous EHs are color-coded under a shadowed background.

pyridine-3,5-dicarboxylate (~1%). Therefore, even though the EHs in this study were identified by a selection process based on the utilization of short esters (see Supporting Information Methods), the isolation of EHs with ample substrate spectra and the ability to hydrolyze very large substrates was not compromised.

We detected drastic shifts in substrate specificity (Figure 2), with glyceryl tripropionate as the only substrate hydrolyzed by all EHs. This is consistent with the high sequence variability within EHs, with an average pairwise identity of 13.74%. We then sought to determine the primary factors shaping the substrate range and thus defined different functional clusters. First, we observed that global sequence identity was of limited relevance for inferring the substrate range because no correlation was found (r^2 = 0.25) between the differences in identity and the number of esters that were hydrolyzed (Supporting Information Tables S1 and S2). Second, comparisons of the substrate range and the hydrolysis rate (U g^{-1} for the best substrates) were performed (Supporting Information Table S1). No correlation existed (r^2 = 0.073), suggesting that our assay conditions allow evaluating the promiscuity level whatever the hydrolytic rate of the EH is. In addition to the low correlation values, no threshold above or below which one could qualitatively classify the substrate range was observed in both cases, so that sequence identity and hydrolytic rate are neither predictive nor classification parameters of promiscuity. Additionally, no link between substrate range and habitat was found because EHs from the same biosource fell into separate clusters (Figure 2). Phylogeny-substrate spectrum relationships were further examined. Figure 2 indicates that the broad substrate-spectrum EHs did not cluster in a single phylogenetic branch, yet substrate promiscuity was mostly found for members of one of 10 subfamilies covered. Indeed, 67% of the EHs that could hydrolyze 30 or more esters

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

(mostly located in clusters C1 and C2 in Figure 2) were assigned to FIV,[14,23] and this percentage increased to 84% when considering only those EHs that could hydrolyze 42 to 72 esters (Figure 2; cluster C1). In addition to FIV members, a FVIII serine beta-lactamase showed prominent substrate spectra (see cluster 1). Members of both families (FIV, 8; FVIII, 1; see cluster C1) hydrolyzed as many esters (from 61 to 72) as the yeast family member CalB (68 esters), the most promiscuous commercially available lipase preparation used for the production of fine chemicals.[29]

Phylogeny was thus indicated as a predictive marker of the substrate range of EHs, as although a broad substrate scope was assigned to several sequence clusters, this feature was prevalent in members of FIV. A query sequence that matched FIV could be easily identified by means of the consensus motif GDSAGG around the catalytic serine; this family is also called the hormone-sensitive lipase (HSL) family because a number of FIV EHs display a striking similarity to the mammalian HSL.[14,23] Noticeably, the location of some FIV members in functional clusters with narrow substrate spectra (Figure 2) suggests that factors other than phylogeny contribute to the substrate spectra of EHs.

The Active Site Effective Volume Is a Prominent Marker of EH Promiscuity. Structural-to-substrate spectrum relationships were further examined by protein−ligand simulations to find additional markers of promiscuity. Crystals from recombinant EH1,[28] the protein with the broadest substrate range under our assay conditions, were obtained as described in the Supporting Information Methods. The enzyme with the widest substrate range was considered the best candidate for understanding the nature of promiscuity. This enzyme seems to have a wide active site environment as, under our assay conditions, it accepted 72 esters ranging from short (*e.g.*, vinyl acetate) to large (*e.g.*, 2,4-dichlorobenzyl-2,4-dichlorobenzoate; Figure 2). We also obtained crystals of recombinant EH102, which was isolated from the same habitat[28] but had a restricted substrate range, hydrolyzing only 10 of the 96 esters tested (Figure 2). Crystallographic data and refinement statistics for the two structures are given in Supporting Information Table S4.

To rationalize the substrate range shown by EH1 and EH102, we performed substrate migration studies using the software Protein Energy Landscape Exploration (PELE), which is an excellent tool to map ligand migration and binding, as shown in studies with diverse applications.[30−32] To map the tendency of a substrate to remain close to the catalytic triad, the substrate was placed in a catalytic position, within a proton abstraction distance from the catalytic serine, and allowed to freely explore the exit from the active site. The PELE results for both proteins and glyceryl triacetate are shown in Figure 3a. Clearly, EH1 has a significantly better binding profile, with an overall lower binding energy and a better funnel shape, whereas EH102 had a qualitatively unproductive binding-energy profile. This difference in the binding mechanism can be explained by the catalytic triad environment. EH1 has a somewhat wide but buried active site, whereas EH102 has a surface-exposed catalytic triad (Figure 4a). These structural differences translate into significant changes in the active site volume, as defined using Fpocket; the active site cavity of EH1 is 3-fold larger than that of EH102. Moreover, important changes are observed when inspecting the solvent exposure of the cavity. Figure 3b shows the relative solvent accessible surface area (SASA) for the substrate along the exploration of PELE, computed as a (dimensionless) percentage (0−1) of the ligand SASA in solution. Even at catalytic positions (distance Ser(O)−substrate(C) ∼ 3−4 Å), in EH102 we observe that ∼40% of the surface of the substrate is accessible to the solvent, which greatly destabilizes the substrate and facilitates escape to the bulk solvent. By contrast, EH1 has a larger but almost fully occluded site, with relative SASA values of approximately 0−10%, which can better stabilize the substrate.

After defining key points underlying the promiscuity of EH1, i.e., a larger active site volume and a lower SASA (Figure 4a), we extended the analysis to other EHs. First, we collected all 11 available crystal structures (Supporting Information Table S1) and computed the active site volume and relative SASA of the catalytic triad (Figure 5, square symbols). We next extended the analysis to the rest of the EHs using homology modeling (using the 11 crystals available) and produced a structural model for 84 additional enzymes. The missing ones were those with sequence identities of less than 25% (to an existing crystal) or those for which the catalytic triad could not be unambiguously identified (*i.e.*, not suitable alignments). Figure 5 (circle symbols) shows the active site effective volume data for all structural models. The analysis indicated a ratio threshold of 62.5 Å^3 for qualitatively classifying substrate promiscuity. Note that the relative SASA of the catalytic triad (derived from the GetArea server, see Supporting Information Methods) adopts values of 0−100; the actual value of the effective volume threshold will depend on the chosen range. We observed that values equal to or higher

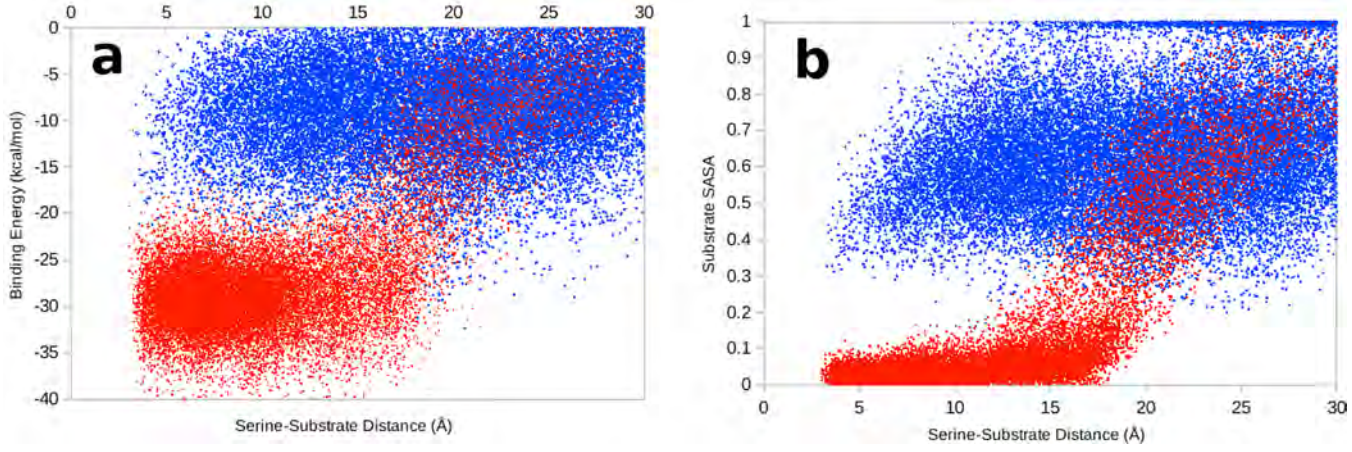


Figure 3. Protein Energy Landscape Exploration (PELE) analysis. Panel a shows the protein−substrate interaction plots for EH1 (red) and EH102 (blue). Panel b shows the relative SASA for glyceryl triacetate in EH1 (red) and EH102 (blue) computed as a dimensionless ratio (0−1) using PELE.

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

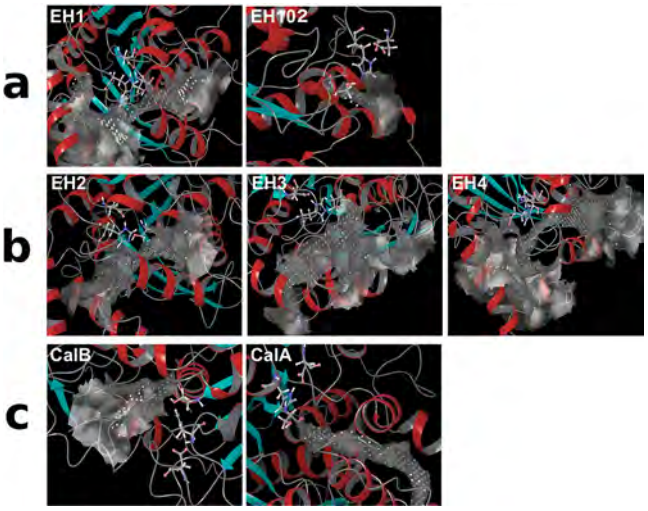


Figure 4. Catalytic triad exposure of selected EHs with the broadest and lowest substrate ranges. (a) The catalytic triad (ball-and-sticks) and the main adjacent cavity (gray clouds) as detected by SiteMap are underlined to demonstrate the differences between a promiscuous (EH1) and nonpromiscuous (EH102) EHs. EH1 can hydrolyze 72 esters and has a defined hidden binding cavity (effective volume: 166.7 Å^3). EH102, by contrast, can hydrolyze only 10 esters and has a surface-exposed triad (high SASA) and an almost negligible binding cavity (38.5 Å^3). The three top EHs with the broadest substrate ranges (b), positioned in the ranking after EH1, and the commercial CalB and CalA lipases (c), are also represented. On each panel, we highlight the catalytic triad and the main adjacent cavity as detected by SiteMap, demonstrating the differences in active site topology. EH2, EH3, and EH4, all assigned to FIV (as EH1), hydrolyzed 71, 69, and 67 esters and have defined but distinct hidden binding cavities (500, 200, and 200 Å^3, in the same order), as EH1. CalB, which was capable of hydrolyzing 68 esters, has a binding cavity (200 Å^3) that is also hidden but highly different from those of the other EHs. CalA, by contrast, hydrolyzed only 36 esters and has a low surface-exposed triad (SASA), with restrictive access to the catalytic triad (1000 Å^3).

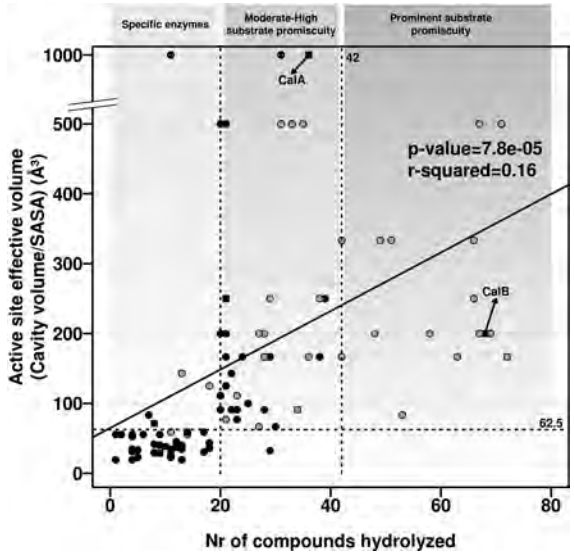


Figure 5. Defining of the substrate range of the EH by topology of the catalytic environment. The figure shows the relationships between the active site effective volume (in Å^3) and enzyme promiscuity (number of substrates hydrolyzed). Note that the presented data were obtained using the active site cavity volume computed in Å^3 and SASA as a dimensionless ratio from 0 to 100 using the GetArea server (http://curie.utmb.edu/getarea.html). The panel contains information for EHs for which crystal structures (square) and homology models (circles) could be unambiguously established (sequence identity ≥25%) and the catalytic triad identified. Gray circles and squares indicate the EHs assigned to FIV. The analysis indicated a threshold ratio (indicated by a horizontal dashed gray line) at which it is possible to qualitatively classify substrate promiscuity based on hydrolysis of at least 20 substrates. Phylogenetic analysis further extended the substrate spectra to ≥42 esters, as only enzymes assigned to FIV and conforming to the 62.5 Å^3 threshold, together with CalB, were capable of converting such a high number of esters. The positioning for the commercial CalA and CalB lipases is indicated.

than 62.5 Å^3 corresponded to EHs with activity for 20 or more of the 96 substrates tested and the opposite. There were only six outliers out of 95 EHs that did not follow this rule. Thus, the performance is of excellent (with 94%) accuracy if used as a classifier. The effective volume, however, does not have quantitative predictions for the exact number of esters hydrolyzed (r^2 = 0.16 for data in Figure 5), most likely because above the 62.5 Å^3 threshold, the capability to hydrolyze more or less substrates may specifically depend on the topology of the catalytic environment (Figure 4a–c), which may differ within families. Particularly, none of the different family members that conformed to the ≥62.5 Å^3 threshold, except those from FIV (*i.e.*, at least 50% of its members as shown in Figure 5, gray circle symbols) and CalB, could hydrolyze 42 or more esters. Therefore, the classification potential of the effective volume measure increased when combined with phylogenetic data. Noticeably, we observed that the predictive capacity of cavity volume/SASA is not influenced by the presence of flexible elements in the structure (Supporting Information Results).

The Active Site Effective Volume Is Also Indicative of Molecules Being Accepted As Substrates. We further used the active site cavity volume/SASA to also dissect its role in substrate specificity. We restricted the analysis to the 96 EHs for which this value could be unambiguously calculated (see above). The analysis indicated that the conversion of 34 esters was only observed for EHs conforming to the ≥62.5 Å^3 threshold (Supporting Information Figure S3). All but two (vinyl crotonate and ethyl acetate) could be considered large alkyl or hindered aromatic esters and included important molecules in synthetic organic chemistry such as paraben esters. This suggests that active sites with larger volume and a lower SASA (i.e., cavity less exposed to the surface) will most likely support hydrolysis of these esters. Therefore, the effective volume measure could be used to some extent as an indicator of substrates that may or may not be hydrolyzed by EHs. However, not all EHs fitting the ≥62.5 Å^3 threshold could convert all 34 of these esters, implying that this measure does not allow deepening into substrate specificity, which may depend on the topology of the catalytic environments as mentioned previously (Figure 4a–c). However, we found that the probability that benzyl-, butyl-, and propyl-paraben esters, major intermediates in chemical synthesis, are converted by members of the FIV with an effective volume ≥ 62.5 Å^3 is significantly higher (~35%) than that of EHs from FIV with a volume < 62.5 Å^3 and EHs from other families, whatever the value of the effective volume (approaching zero percent in our study); for those EHs for which effective volume could not be calculated, this probability is as low as 1.9% (Supporting Information Figure S4). This again exemplifies that the effective volume measure, when combined with phylogenetic information, is not only indicative of a promiscuity level but also can be used to predict the capacity to hydrolyze esters such as paraben esters. Screen programs to find EHs capable of converting

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225–234

paraben esters should most likely be directed to find those assigned to FIV and with cavity volume/SASA ≥ 62.5 Å^3.

The Effective Volume Is Also a Marker of Substrate Promiscuity in Proteins Other than EHs. In order to evaluate the possibility that the active site effective volume may be a marker of substrate promiscuity in other enzymes, substrate spectra-effective volume relationships should be investigated in other protein families. In this line, Huang *et al.*[1] recently performed a systematic analysis of the substrate spectra of 200 phosphatases of the HAD superfamily, when tested against a set of 167 substrates. We collected the available crystal structures of each of the HAD phosphatases (Supporting Information Table S5) and computed the active site effective volume. We restricted the analysis to C2 cap members as they were reported to have a broader substrate spectrum,[1] and crystal structures with low to high effective volume are available. Interestingly, we observed that the effective volume (using the two conserved aspartic catalytic residues as the corrective SASA factor) was highly correlated (r^2 = 0.92) with the substrate range (Figure 6). Thus, the effective volume can be used as a molecular

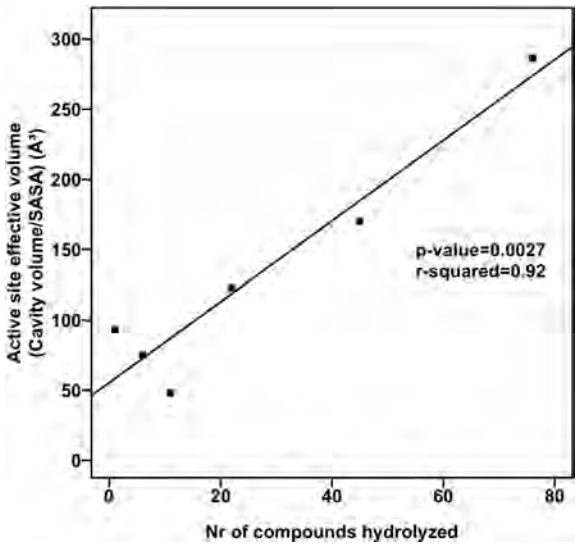


Figure 6. Relationships between the active site effective volume (in Å^3) and enzyme promiscuity (number of substrates hydrolyzed) of C2 members of HAD phosphatases. The number of substrates converted by each HAD phosphatase was obtained from Huang *et al.*[1] and is summarized in Supporting Information Table S5. The panels contain information for HAD phosphatases for which crystal structures were available and the catalytic residues identified. The active site effective volume (in Å^3) was calculated as described in Figure 5.

classification parameter of substrate promiscuity of phosphatases of the HAD superfamily when crystal structures are available. When this analysis was extended to the rest of the enzymes using homology modeling, we observed a similar trend to that of EHs (Supporting Information Figure S5). That is, no correlation existed (r^2 = 0.043), but still the effective volume can be used as a classifier of the substrate range as for EHs. Indeed, although a threshold could not be unambiguously established, sequences with the top 10 effective volumes belong to moderate-high to high promiscuity enzymes.

In conclusion, we found that the topology around the catalytic position, by means of an active site effective volume (cavity volume/SASA) threshold, is a dominant criterion of substrate promiscuity in EHs, which can be further extended by adding phylogenetic analysis. The rationale behind this parameter is as follows. Large volumes increase promiscuity until a certain value at which the cavity becomes too exposed and is not capable of properly accommodating and, importantly, retaining the substrate in specific catalytic binding interactions. This point is well captured by the SASA percentage of the catalytic triad, a dimensionless ratio that corrects for large volume measures in exposed sites. Importantly, the parameters of active site volume and relative SASA can be easily transferred to other systems. Indeed, the fact that the EHs investigated herein have different folds and that this parameter was also a marker of substrate spectra for phosphatases of the HAD superfamily opens the possibility of applying the effective volume measure to other enzymes requiring substrate anchoring. In all cases, the effective volume threshold-to-substrate relationships must be established. We would like to make note that the active site volume is not a static property, as the active site will breathe, depending on how flexible the protein is. In addition to that, the 62.5 Å^3 threshold for qualitatively classifying substrate promiscuity is based on the analysis of 147 EHs when tested against 96 esters. Although increasing the number of EHs and esters may influence this threshold and increase accuracy, it will not affect the fact that the measurement of the effective volume (cavity volume/SASA) can be used as the first molecular classification method of substrate promiscuity in EHs.

Our measurement is not a quantitative one, but rather a qualitative ranking (classification) procedure that will allow, for example, selecting sequences in databases for expression, particularly those encoding promiscuous enzymes capable of converting multiple substrates. This will substantially reduce reagent and labor costs compared to methods requiring the extensive cloning of all genes, and the expression and characterization of all enzymes in databases to later find those being promiscuous.[33] This possibility was herein examined by successfully mapping the open reading frames from the TARA Oceans project assemblies[34] and by identifying a high number of sequences encoding EHs with presumptive prominent substrate promiscuity (Supporting Information Results, Figures S6 and S7). Application of the effective volume measure to examine the sequences daily generated or deposited in databases requires having some crystals or X-ray structures for the model production. This limitation prevents predicting promiscuity from sequences lacking any structural information. Indeed, 36% of the EHs in this study (52 of the 147, including CalA/B) could not be included in the correlation because no calculation was possible. Accumulation of structural information and design and application of better modeling algorithms in the future will help solving this limitation.[35] Future studies might also explore molecular dynamics (MD) simulations to measure also the flexibility of the active site and not just the size of the cavity. By using this strategy, it was recently reported that the broad promiscuity of the members of the alkaline phosphatase superfamily arises from cooperative electrostatic interactions in the active site, allowing each enzyme to adapt to the electrostatic needs of different substrates.[36] In the particular case of EH phylogeny, a marker which does not require a three-dimensional structure was also suggested as a predictive classification marker of the substrate range. Indeed, this study suggests that in case of an unknown EH for which a crystal structure is not available or a homology model could not be established, then its assignation to family IV[14,23] increases the likelihood that this EH is promiscuous.

The present study not only provides clear evidence that substrate promiscuity in EHs has evolved from different core structural domains fitting an effective volume around the active

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

site, albeit with a bias toward that occurring in FIV members, but also from different phylogenetic lineages, many of which remain underexplored to date (Supporting Information Results and Figure S2). These are new findings as it was previously thought that the substrate range in a superfamily increased from a single ancestral core domain,[1] and because the identities of some microbial groups containing promiscuous enzymes, herein EHs, were previously unknown. Finally, this study also enabled the selection of a set of EH candidates that can compete with best commercial EHs such as CalB, as they show a broader substrate profile and specific activities up to 3-fold higher (Supporting Information Table S6). Their sequences can be used to search databases for similar promiscuous EHs. Further investigations should also determine the occurrence of other types of promiscuous EH phenotypes with broader substrate ranges than those identified in this study. For example, at least the stability of substrate-promiscuous EHs at different temperatures and with various solvents, along with the occurrence and evolution of secondary reactions, should be investigated in terms of condition and catalytic promiscuity.

METHODS

Protein Samples. Two main sources of EHs were used in the present study, all of them isolated *via* naïve and sequence-based screens in genomes and metagenomes. A first set of samples was EHs previously reported, as in the bibliography (69 in total), and that were herein substrate-profiled for first time. A second set was EHs (77) that are herein reported for first time. The extensive details of the source, cloning, expression, and purification of each of the active and soluble EHs are provided in the Supporting Information Methods and Table S1.

Ester Bond Hydrolysis Activity Assessment: Substrate Profiling Tests with 96 Esters. Hydrolytic activity was assayed at 550 nm using 96 structurally diverse esters in 384-well plates as previously described.[24,27,28] Before the assay, a concentrated stock solution of the esters was prepared at a concentration of 100 mg mL^{-1} in acetonitrile and dimethyl sulfoxide (DMSO). The assays were conducted according to the following steps. First, a 384-well plate (Molecular Devices, LLC, CA, USA) was filled with 20 μL of 5 mM *N*-(2-hydroxyethyl)piperazine-*N*′-(3-propanesulfonic acid (EPPS) buffer, at pH 8.0, using a QFill3 microplate filler (Molecular Devices, LLC, CA, USA). Second, 2 μL of each ester stock solution was added to each well using a PRIMADIAG liquid-handling robot (EYOWN TECHNOLOGIES SL, Madrid, Spain). The ester was dispensed in replicates. After adding the esters, the 384-well plate was filled with 20 μL of 5 mM EPPS buffer, at pH 8.0, containing 0.912 mM Phenol Red (used as a pH indicator) using a QFill3 microplate filler. The final ester concentration of the ester in each well was 1.14 mg mL^{-1}, and the final concentration of Phenol Red was 0.45 mM. A total of 2 μL of protein extract (containing 1−5 mg mL^{-1} pure protein or 200 mg mL^{-1} wet cells expressing proteins) was immediately added to each well using an Eppendorf Repeater M4 pipet (Eppendorf, Hamburg, Germany) or a PRIMADIAG liquid-handling robot. Accordingly, the total reaction volume was 44 μL, with 4.5% (v/v) acetonitrile or DMSO in the reaction mixture. After incubation at 30 °C in a Synergy HT Multi-Mode Microplate Reader, ester hydrolysis was measured spectrophotometrically in continuous mode at 550 nm for a total time of 24 h. Commercially available CALA L and CALB L (Novozymes A/S, Bagsvaerd, Denmark) were diluted 10-fold with 5 mM EPPS buffer, at pH 8.0, and 2 μL of this solution was used immediately for reaction tests under the conditions described before. In all cases, specific activities (in U g^{-1} protein) were determined. One unit (U) of enzyme activity was defined as the amount of wet cells expressing EHs or pure EHs required to transform 1 μmol of substrate in 1 min under the assay conditions using the reported extinction coefficient ($\varepsilon_{\text{Phenol-red}}$ at 550 nm = 8450 M^{-1} cm^{-1}). All values were corrected for nonenzymatic transformation (i.e., the background rate) and for the background signal using *E. coli* cells that did not express any target protein (control cells included empty vectors). Note that a positive reaction was indicated by the restrictive criterion of a change greater than 6-fold above the background signal. Specific activity determinations (in U g^{-1}) for wet cells expressing each of the selected EHs or pure or commercial proteins are available in Supporting Information Tables S3 and S6, respectively.

Structural Determinations and Homology Modeling. The proteins EH1 and EH102 were expressed, purified, and crystallized using the sitting-drop method in Intelliplate 96-well plates and a Mosquito liquid-handling robot (TTP LabTech) according to previously described procedures.[37] For EHs for which crystal structures were not available, homology models were developed using Prime software from Schrödinger. Prime uses BLAST (with BLOSUM62 matrix) for homology search and alignment and refines the results using the Pfam database and pairwise alignment with ClustalW.

Protein Energy Landscape Exploration (PELE) Simulations. We used Protein Energy Landscape Exploration (PELE) software to sample the binding modes of glyceryl triacetate with EH1 and EH102.[38,39] The initial structures were taken from the coordinates of the EH1 and EH102 crystal structures (PDB codes: 5JD4 and 5JD3, respectively). The protonation state of titratable residues was estimated with the Protein Preparation Wizard (PROPKA)[40] and the H++ server (http://biophysics.cs.vt.edu/H++) followed by visible inspection. At pH 8 (the pH at which the activity assays were performed), the catalytic triad histidine residues were δ-protonated, and the catalytic triad aspartic acid residues were deprotonated, resulting in the formation of a histidine-serine and histidine-aspartic hydrogen-bonding network. The glyceryl triacetate structure was fully optimized with Jaguar[41] in an implicit solvent, and the electrostatic potential charges were computed with the density functional M06 at the 6-31G* level of theory. The ligand parameters were extracted from these for the classic simulations.

Cavity Volume and Solvent Accessible Surface Area (SASA) calculation. The relative Solvent Accessible Surface Area (SASA) for a residue was obtained using the GetArea Web server.[42] Cavity volumes were computed with Fpocket,[43] a very fast open-source protein pocket (cavity) detection algorithm based on Voronoi tessellation. Fpocket includes two other programs (dpocket and tpocket) that allow the extraction of pocket descriptors and the testing of owned scoring functions, respectively.

For extensive details of the methods, see Supporting Information Methods.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschembio.7b00996.

Supporting Results, Methods, Figures S1−S7, and Table S4 (PDF)

Tables S1−S3, S5, and S6 (XLS)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: victor.guallar@bsc.es.

*E-mail: mferrer@icp.csic.es.

ORCID

Gerard Santiago: 0000-0002-0506-3049

Jürgen Pleiss: 0000-0003-1045-8202

Alexander F. Yakunin: 0000-0003-0813-6490

Víctor Guallar: 0000-0002-4580-1114

Manuel Ferrer: 0000-0003-4962-4714

Present Addresses

[γ]Current address: School of Chemistry, Bangor University, LL57 2UW Bangor, UK.

[δ]Current address: Lehrstuhl für Biotechnologie, RWTH Aachen University, Aachen, Germany.

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

[e]Current address: Carl R. Woese Institute for Genomic Biology, Urbana, USA.

Author Contributions

[ζ]These authors contributed equally to this work.

Author Contributions

[ψ]These authors contributed equally in coordinating activities.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

C.C. thanks the Spanish Ministry of Economy, Industry and Competitiveness for a Ph.D. fellowship (Grant BES-2015-073829). V.M. thanks the Francisco José de Caldas Scholarship Program (Administrative Department of Science, Technology and Innovation, COLCIENCIAS). The authors acknowledge the members of the MAMBA, MAGICPAH, ULIXES, KILLSPILL and INMARE Consortia for their support in sample collection. David Rojo is also acknowledged for his valuable help with log P calculations. This project received funding from the European Union's Horizon 2020 research and innovation program [Blue Growth: Unlocking the potential of Seas and Oceans] under grant agreement no. 634486 (project acronym INMARE). This research was also supported by the European Community Projects MAGICPAH (FP7-KBBE-2009-245226), ULIXES (FP7-KBBE-2010-266473), and KILLSPILL (FP7-KBBE-2012-312139) and grants BIO2011-25012, PCIN-2014-107, BIO2014-54494-R, and CTQ2016-79138-R from the Spanish Ministry of Economy, Industry and Competitiveness. The present investigation was also funded by the Spanish Ministry of Economy, Industry and Competitiveness within the ERA NET IB2, grant no. ERA-IB-14-030 (MetaCat), the UK Biotechnology and Biological Sciences Research Council (BBSRC), grant no. BB/M029085/1, and the German Research Foundation (FOR1296). R.B. and P.N.G. acknowledge the support of the Supercomputing Wales project, which is part-funded by the European Regional Development Fund (ERDF) *via* the Welsh Government. O.V.G. and P.N.G. acknowledge the support of the Centre of Environmental Biotechnology Project funded by the European Regional Development Fund (ERDF) through the Welsh Government. A.Y. and A.S. gratefully acknowledge funding from Genome Canada (2009-OGI-ABC-1405) and the NSERC Strategic Network grant IBN. A.I.P. was supported by the Counseling of Economy and Employment of the Principality of Asturias, Spain (Grant FC-15-GRUPIN14-107). V.G. acknowledges the joint BSC-CRG-IRB Research Program in Computational Biology. The authors gratefully acknowledge financial support provided by the European Regional Development Fund (ERDF).

■ REFERENCES

(1) Huang, H., Pandya, C., Liu, C., Al-Obaidi, N. F., Wang, M., Zheng, L., Toews Keating, S., Aono, M., Love, J. D., Evans, B., Seidel, R. D., Hillerich, B. S., Garforth, S. J., Almo, S. C., Mariano, P. S., Dunaway-Mariano, D., Allen, K. N., and Farelli, J. D. (2015) Panoramic view of a superfamily of phosphatases through substrate profiling. *Proc. Natl. Acad. Sci. U. S. A. 112*, E1974−1983.

(2) Huang, R., Hippauf, F., Rohrbeck, D., Haustein, M., Wenke, K., Feike, J., Sorrelle, N., Piechulla, B., and Barkman, T. J. (2012) Enzyme functional evolution through improved catalysis of ancestrally non-preferred substrates. *Proc. Natl. Acad. Sci. U. S. A. 109*, 2966−2971.

(3) Yip, S. H., and Matsumura, I. (2013) Substrate ambiguous enzymes within the *Escherichia coli* proteome offer different evolutionary solutions to the same problem. *Mol. Biol. Evol. 30*, 2001−2012.

(4) Price, D. R., and Wilson, A. C. (2014) Substrate ambiguous enzyme facilitates genome reduction in an intracellular symbiont. *BMC Biol. 12*, 110.

(5) Devamani, T., Rauwerdink, A. M., Lunzer, M., Jones, B. J., Mooney, J. L., Tan, M. A., Zhang, Z. J., Xu, J. H., Dean, A. M., and Kazlauskas, R. J. (2016) Catalytic promiscuity of ancestral esterases and hydroxynitrile lyases. *J. Am. Chem. Soc. 138*, 1046−1056.

(6) Hult, K., and Berglund, P. (2007) Enzyme promiscuity: mechanism and applications. *Trends Biotechnol. 25*, 231−238.

(7) Copley, S. D. (2015) An evolutionary biochemist's perspective on promiscuity. *Trends Biochem. Sci. 40*, 72−78.

(8) London, N., Farelli, J. D., Brown, S. D., Liu, C., Huang, H., Korczynska, M., Al-Obaidi, N. F., Babbitt, P. C., Almo, S. C., Allen, K. N., and Shoichet, B. K. (2015) Covalent docking predicts substrates for haloalkanoate dehalogenase superfamily phosphatases. *Biochemistry 54*, 528−537.

(9) Nobeli, I., Favia, A. D., and Thornton, J. M. (2009) Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol. 27*, 157−167.

(10) Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W. R., Golyshina, O. V., and Golyshin, P. N. (2016) Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb. Biotechnol. 9*, 22−34.

(11) Braakman, R., and Smith, E. (2014) Metabolic evolution of a deep-branching hyperthermophilic chemoautotrophic bacterium. *PLoS One 9*, e87950.

(12) Giovannoni, S. J., Cameron Thrash, J., and Temperton, B. (2014) Implications of streamlining theory for microbial ecology. *ISME J. 8*, 1553−1565.

(13) Lan, T., Wang, X. R., and Zeng, Q. Y. (2013) Structural and functional evolution of positively selected sites in pine glutathione S-transferase enzyme family. *J. Biol. Chem. 288*, 24441−24451.

(14) Ferrer, M., Bargiela, R., Martínez-Martínez, M., Mir, J., Koch, R., Golyshina, O. V., and Golyshin, P. N. (2015) Biodiversity for biocatalysis: A review of the α/β-hydrolase fold superfamily of esterases-lipases discovered in metagenomes. *Biocatal. Biotransform. 33*, 235−249.

(15) Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M., and Witholt, B. (2001) Industrial biocatalysis today and tomorrow. *Nature 409*, 258−268.

(16) Koudelakova, T., Chovancova, E., Brezovsky, J., Monincova, M., Fortova, A., Jarkovsky, J., and Damborsky, J. (2011) Substrate specificity of haloalkane dehalogenases. *Biochem. J. 435*, 345−354.

(17) Risso, V. A., Gavira, J. A., Mejia-Carmona, D. F., Gaucher, E. A., and Sanchez-Ruiz, J. M. (2013) Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases. *J. Am. Chem. Soc. 135*, 2899−2902.

(18) Amin, S. R., Erdin, S., Ward, R. M., Lua, R. C., and Lichtarge, O. (2013) Prediction and experimental validation of enzyme substrate specificity in protein structures. *Proc. Natl. Acad. Sci. U. S. A. 110*, E4195−4202.

(19) Anton, B. P., Chang, Y. C., Brown, P., Choi, H. P., Faller, L. L., Guleria, J., Hu, Z., Klitgord, N., Levy-Moonshine, A., Maksad, A., Mazumdar, V., McGettrick, M., Osmani, L., Pokrzywa, R., Rachlin, J., Swaminathan, R., Allen, B., Housman, G., Monahan, C., Rochussen, K., Tao, K., Bhagwat, A. S., Brenner, S. E., Columbus, L., de Crécy-Lagard, V., Ferguson, D., Fomenkov, A., Gadda, G., Morgan, R. D., Osterman, A. L., Rodionov, D. A., Rodionova, I. A., Rudd, K. E., Söll, D., Spain, J., Xu, S. Y., Bateman, A., Blumenthal, R. M., Bollinger, J. M., Chang, W. S., Ferrer, M., Friedberg, I., Galperin, M. Y., et al. (2013) The COMBREX project: design, methodology, and initial results. *PLoS Biol. 11*, e1001638.

(20) Colin, P. Y., Kintses, B., Gielen, F., Miton, C. M., Fischer, G., Mohamed, M. F., Hyvönen, M., Morgavi, D. P., Janssen, D. B., and Hollfelder, F. (2015) Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nat. Commun. 6*, 10008.

(21) Chen, C., Huang, H., and Wu, C. H. (2017) Protein bioinformatics databases and resources. *Methods Mol. Biol. 1558*, 3−39.

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

(22) Fischer, M., and Pleiss, J. (2003) The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Res. 31*, 319−321.
(23) Arpigny, J. L., and Jaeger, K. E. (1999) Bacterial lipolytic enzymes: classification and properties. *Biochem. J. 343*, 177−183.
(24) Alcaide, M., Tornés, J., Stogios, P. J., Xu, X., Gertler, C., Di Leo, R., Bargiela, R., Lafraya, A., Guazzaroni, M. E., López-Cortés, N., Chernikova, T. N., Golyshina, O. V., Nechitaylo, T. Y., Plumeier, I., Pieper, D. H., Yakimov, M. M., Savchenko, A., Golyshin, P. N., and Ferrer, M. (2013) Single residues dictate the co-evolution of dual esterases: MCP hydrolases from the α/β hydrolase family. *Biochem. J. 454*, 157−166.
(25) Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P. M., and Henrissat, B. (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem. J. 432*, 437−444.
(26) Popovic, A., Hai, T., Tchigvintsev, A., Hajighasemi, M., Nocek, B., Khusnutdinova, A. N., Brown, G., Glinos, J., Flick, R., Skarina, T., Chernikova, T. N., Yim, V., Brüls, T., Paslier, D. L., Yakimov, M. M., Joachimiak, A., Ferrer, M., Golyshina, O. V., Savchenko, A., Golyshin, P. N., and Yakunin, A. F. (2017) Activity screening of environmental metagenomic libraries reveals novel carboxylesterase families. *Sci. Rep. 7*, 44103.
(27) Janes, L. E., Löwendahl, C., and Kazlauskas, R. J. (1998) Rapid quantitative screening of hydrolases using pH indicators. Finding enantioselective hydrolases. *Chem. Eur. J. 4*, 2317−2324.
(28) Martínez-Martínez, M., Alcaide, M., Tchigvintsev, A., Reva, O., Polaina, J., Bargiela, R., Guazzaroni, M. E., Chicote, A., Canet, A., Valero, F., Rico Eguizabal, E., Guerrero, Mdel C., Yakunin, A. F., and Ferrer, M. (2013) Biochemical diversity of carboxyl esterases and lipases from Lake Arreo (Spain): a metagenomic approach. *Appl. Environ. Microbiol. 79*, 3553−3562.
(29) Daiha, K. d. G., Angeli, R., de Oliveira, S. D., and Almeida, R. V. (2015) Are lipases still important biocatalysts? A study of scientific publications and patents for technological forecasting. *PLoS One 10*, e0131624.
(30) Borrelli, K. W., Cossins, B., and Guallar, V. (2009) Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility. *J. Comput. Chem. 31*, 1224−1235.
(31) Hernández-Ortega, A., Borrelli, K., Ferreira, P., Medina, M., Martínez, A. T., and Guallar, V. (2011) Substrate diffusion and oxidation in GMC oxidoreductases: an experimental and computational study on fungal aryl-alcohol oxidase. *Biochem. J. 436*, 341−350.
(32) Santiago, G., de Salas, F., Lucas, F., Monza, E., Acebes, S., Martínez, A., Camarero, S., and Guallar, V. (2016) Computer-aided laccase engineering: toward biological oxidation of arylamines. *ACS Catal. 6*, 5415−5423.
(33) Barak, Y., Nov, Y., Ackerley, D. F., and Matin, A. (2008) Enzyme improvement in the absence of structural knowledge: a novel statistical approach. *ISME J. 2*, 171−179.
(34) Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., d'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., Kokoszka, F., Lepoivre, C., Lima-Mendez, G., Poulain, J., Poulos, B. T., Royo-Llonch, M., Sarmento, H., Vieira-Silva, S., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Bowler, C., de Vargas, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Jaillon, O., Not, F., Ogata, H., Pesant, S., Speich, S., Stemmann, L., Sullivan, M. B., Weissenbach, J., Wincker, P., Karsenti, E., Raes, J., Acinas, S. G., Bork, P., et al. (2015) Structure and function of the global ocean microbiome. *Science 348*, 1261359.
(35) Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., and Tramontano, A. (2016) Critical assessment of methods of protein structure prediction: Progress and new directions in round XI. *Proteins: Struct., Funct., Genet. 84* (Suppl 1), 4−14.
(36) Barrozo, A., Duarte, F., Bauer, P., Carvalho, A. T. P., and Kamerlin, S. C. L. (2015) Cooperative electrostatic interactions drive functional evolution in the alkaline phosphatase superfamily. *J. Am. Chem. Soc. 137*, 9061−9076.
(37) Alcaide, M., Stogios, P. J., Lafraya, Á., Tchigvintsev, A., Flick, R., Bargiela, R., Chernikova, T. N., Reva, O. N., Hai, T., Leggewie, C. C., Katzke, N., La Cono, V., Matesanz, R., Jebbar, M., Jaeger, K. E., Yakimov, M. M., Yakunin, A. F., Golyshin, P. N., Golyshina, O. V., Savchenko, A., and Ferrer, M. (2015) Pressure adaptation is linked to thermal adaptation in salt-saturated marine habitats. *Environ. Microbiol. 17*, 332−345.
(38) Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B 105*, 6474−6487.
(39) Borrelli, K. W., Vitalis, A., Alcantara, R., and Guallar, V. (2005) PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J. Chem. Theory Comput. 1*, 1304−1311.
(40) Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput.-Aided Mol. Des. 27*, 221−234.
(41) Bochevarov, A. D., Harder, E., Hughes, T. F., Greenwood, J. R., Braden, D. A., Philipp, D. M., Rinaldo, D., Halls, M. D., Zhang, J., and Friesner, R. A. (2013) Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem. 113*, 2110−2142.
(42) Fraczkiewicz, R., and Braun, W. (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem. 19*, 319.
(43) Le Guilloux, V., Schmidtke, P., and Tuffery, P. (2009) Fpocket: An open source platform for ligand pocket detection. *BMC Bioinf. 10*, 168.

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234