Discovering novel hydrolases from hot environments

Wohlgemuth, Roland; Littlechild, Jennifer; Monti, Daniela; Schnorr, Kirk; van Rossum, Teunke; Siebers, Bettina; Menzel, Peter; Kublanov, Ilya V.; Rike, Anne Gunn; More Authors

DOI
10.1016/j.biotechadv.2018.09.004

Publication date
2018

Document Version
Final published version

Published in
Biotechnology Advances

Biotechnology Advances 36 (2018) 2077–2100

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Discovering novel hydrolases from hot environments

Roland Wohlgemuth[a,o,*], Jennifer Littlechild[b], Daniela Monti[c], Kirk Schnorr[d], Teunke van Rossum[e,p], Bettina Siebers[f], Peter Menzel[g,m], Ilya V. Kublanov[h], Anne Gunn Rike[i,q], Georgios Skretas[j], Zalan Szabo[k,n], Xu Peng[g], Mark J. Young[l]

[a] Sigma-Aldrich, Member of Merck Group, Industriestrasse 25, CH-9470 Buchs, Switzerland
[b] Biocatalysis Centre, University of Exeter, The Henry Wellcome Building for Biocatalysis, Stocker Road, Exeter EX4 4QD, United Kingdom
[c] Istituto di Chimica del Riconoscimento Molecolare, CNR, Via Mario Bianco 9, 20131 Milano, Italy
[d] Novozymes A/S, Krogshoejvej 36, 2880 Bagsvaerd, Denmark
[e] Wageningen University and Research, Laboratory of Microbiology, Stippeneng 4, 6708 WE Wageningen, the Netherlands
[f] Molecular Enzyme Technology and Biochemistry, Biofilm Centre, Faculty of Chemistry, University of Duisburg-Essen, Universitaetsstr. 5, S05 V03 F41, 45141 Essen, Germany
[g] Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 København, Denmark
[h] Extremophiles Metabolism Laboratory, Winogradsky Institute of Microbiology, Research Center of Biotechnology, RAS, Prospekt 60 Letiya Oktyabrya 7, Moscow, Russia
[i] Norwegian Geotechnical Institute, Sognsveien 72, N-0855 Oslo, Norway
[j] Institute of Biology, Medicinal Chemistry & Biotechnology, National Hellenic Research Foundation, 48 Vassileos Constantinou Ave, 11635 Athens, Greece
[k] MicroDish BV, Padualaan 8, 3584 CH Utrecht, the Netherlands
[l] Thermal Biology Institute, Departments of Microbiology and Plant Sciences, 111A Chemistry, Montana State University, Bozeman, MT 59717, USA
[m] Max Delbrück Center for Molecular Medicine, Robert-Rössle-Straße 10, 13125 Berlin, Germany
[n] U-Protein Express BV, Yalelaan 62, 3584 CH Utrecht, the Netherlands
[o] Institute of Technical Biochemistry, Lodz University of Technology, ul. Stefanowskiego 4/10, 90-924 Lodz, Poland
[p] Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, Netherlands
[q] Standard Norge, Mustads vei 1, 0283 Oslo, Norway

ARTICLE INFO

Keywords:
Hydrolases
Metagenomics
Enrichment
Sequencing
Genomics
Thermophiles
Gene expression
Enzyme screening
Enzyme characterization
Biocatalysis

ABSTRACT

Novel hydrolases from hot and other extreme environments showing appropriate performance and/or novel functionalities and new approaches for their systematic screening are of great interest for developing new processes, for improving safety, health and environment issues. Existing processes could benefit as well from their properties. The workflow, based on the HotZyme project, describes a multitude of technologies and their integration from discovery to application, providing new tools for discovering, identifying and characterizing more novel thermostable hydrolases with desired functions from hot terrestrial and marine environments. To this end, hot springs worldwide were mined, resulting in hundreds of environmental samples and thousands of enrichment cultures growing on polymeric substrates of industrial interest. Using high-throughput sequencing and bioinformatics, 15 hot spring metagenomes, as well as several sequenced isolate genomes and transcriptomes were obtained. To facilitate the discovery of novel hydrolases, the annotation platform Anastasia and a whole-cell bioreporter-based functional screening method were developed. Sequence-based screening and functional screening together resulted in about 100 potentially new hydrolases of which more than a dozen have been characterized comprehensively from a biochemical and structural perspective. The characterized hydrolases include thermostable carboxylesterases, enol lactonases, quorum sensing lactonases, gluconolactonases, epoxide hydrolases, and cellulases. Apart from these novel thermostable hydrolases, the project generated an enormous amount of samples and data, thereby allowing the future discovery of even more novel enzymes.

1. Introduction

The search for thermostable hydrolases with novel functions and appropriate performance is of much interest despite the fact that hydrolases have already been workhorses in research and industrial applications (Bornscheuer and Kazlauskas, 2006; Ghisalba et al., 2010;

* Corresponding author at: Institute of Technical Biochemistry, Lodz University of Technology, ul. Stefanowskiego 4/10, 90-924 Lodz, Poland; Sigma-Aldrich, Member of Merck Group, Industriestrasse 25, CH-9470 Buchs, Switzerland.
E-mail addresses: roland.wohlgemuth@p.lodz.pl, roland.wohlgemuth@sial.com (R. Wohlgemuth).

https://doi.org/10.1016/j.biotechadv.2018.09.004
Received 30 May 2018; Received in revised form 21 September 2018; Accepted 24 September 2018
Available online 25 September 2018
0734-9750/

Wohlgemuth, 2010; Bornscheuer et al., 2012; Meyer et al., 2013; Elleuche et al., 2015). The fact that hydrolases have already been well established in large-scale industrial manufacturing environments facilitates further extensions of industrial applications. Thermostable enzymes provide additional benefits such as ease of purification, higher stability, better productivity, enlarged process windows and conditions for more favourable thermodynamics such as a shift of reaction equilibrium (Siddiqui, 2015). These benefits make the identification and characterization of hydrolases from hot terrestrial environments attractive for developing various novel and improved processes in industry. Challenges for discovering hydrolases with the desired properties range from substrate availability, microbial sampling from the hot environments, gene sequencing, bioinformatics, gene expression, to hydrolase characterization and production. This clearly requires a multi-disciplinary approach that combines expertise in sampling and 'in situ' enrichment, metagenomic screening methods, bioinformatics, and high-throughput screening technologies in order to identify novel hydrolases.

Even though the number of sequenced microbial genomes has been rising rapidly in recent years (Land et al., 2015), a large fraction of the global microbial biodiversity remains unexplored. Culture-independent metagenomic sequencing from environmental samples can shed light on this microbial dark matter (for example Rinke et al., 2013; Brown et al., 2015) and facilitate the discovery of novel microbial lineages, especially from extreme environments (Eloe-Fadrosh et al., 2016). Recent advances in computational methods and increasing availability of metagenomic sequencing data enable the assembly of near-complete novel genomes (for example Nielsen et al., 2014; Sangwan et al., 2016). Thus, metagenomics is valuable for searching and discovering novel enzymes from previously uncultured or unknown species.

In this review, we will use HotZyme as an example to describe how novel hydrolases could be screened and discovered. HotZyme is an EU FP7 funded project, aiming to discover novel thermostable hydrolases from hot environments. Targeted enzyme classes include glycosidases, esterases, lactonases, epoxide hydrolases and proteases. A brief summary of the HotZyme workflow is depicted in Fig. 1. Basically, the idea of the project was to explore the biodiversity of hot environments using different mining approaches, involving development of novel bioinformatic tools and platforms and different screening methodologies. In the end, enzymes were subjected to more detailed functional and structural characterization and highly promising enzymes were patented for future potential commercialization. The HotZyme project was started in 2011 and finished in 2015. During this period the HotZyme team collected from hot spring environments globally hundreds of samples, from which thousands of enrichment cultures based on a set of polymeric substrates of industrial interests were performed. Eventually, the team obtained a few dozens of isolates capable of degrading polymeric substrates. Using next-generation sequencing technologies the team sequenced 15 hot spring metagenomes and several isolate genomes and transcriptomes. From the metagenomic libraries, the team identified hundreds of potentially new hydrolases using in silico methods and/or functional screening methods, more than a dozen of which were selected for detailed biochemical and structural analyses.

2. Mining the environment

There were two major objectives for sampling diverse high temperature natural environments found around the world. The first was to obtain environmental DNA samples from the microbes and viruses that inhabit these environments in order to create functional expression libraries and for detailed bioinformatic analysis in search of selected thermostable enzymes. The second objective was to create enrichment cultures of microbes and their viruses for direct screening for selected enzymatic activities. More than three hundred samples were obtained from natural thermal environments located worldwide: terrestrial hot springs of Iceland, Italy, China, Yellowstone National Park (USA), Kamchatka, Kuril Islands, Baikal Lake area (Russia) and deep subsurface biosphere (Western Siberia, Norwegian Sea, Troll, Barents Sea, Spitzbergen). Some representative pictures of the collection sites are shown in Fig. 2. The temperature of the samples was ranging from 40 to 151 °C, while pH varied from 2.0 to 10.5. Samples were represented by water, sediments and microbial biofilms and were subjected to DNA isolation as well as enrichment and isolation of thermophilic microorganisms.

Years of experience of sampling in high temperature environments was brought to the HotZyme project (Bolduc et al., 2012; Inskeep et al., 2013; Inskeep et al., 2010; Bolduc et al., 2015; Colman et al., 2016; Gudbergsdóttir et al., 2016). While it is still challenging to collect and extract sufficient quantity and quality of biomolecules from high temperature environments due to the typically low cell density (as low as 10^2 cells/ml) and geochemical properties of hot spring environments (e.g. low pH and high metal content), it is certainly possible and should not deter researchers new to the field from pursuing field sampling from thermal environments.

2.1. In situ enrichment of thermophilic microorganisms with hydrolytic activities

In situ enrichments with diverse biopolymeric substrates of interest were done in the hot springs of Kamchatka, Kuril Islands and Iceland in a similar way to Kublanov et al. (2009) and Gavrilov et al. (2016), using cellulose (microcrystalline cellulose (MCC), carboxymethyl cellulose (CMC), leaves of corn and bamboo), xylan, starch, alpha- and beta-keratins, xanthan, polyester and polyvinyl alcohol as substrates. The composition of microbial communities developing in primary in situ enrichments was eventually studied by total 16S rRNA gene fragments sequencing using an Illumina platform. Those showing visible degradation of insoluble substrates and/or containing new phylogenetic groups of thermophilic microorganisms were used for further characterization and isolation work.

2.2. High-throughput enrichment and isolation of new thermophilic microorganisms with hydrolytic activities

The aim of this work was to find thermophiles capable of growing on various large molecular weight substrates and thereby to find thermostable hydrolases degrading the primary substrate and ideally also catalyzing the hydrolysis of intermediates. Hundreds of crude environmental samples and in situ enrichments were cultured in the laboratory with media containing different polymeric substrates of industrial interest. The substrates included cellulose, xylan, lignin, starch, chitin, bamboo leaves, α- and β-keratins, xanthan, lichenan, agarose, polyethylene terephthalate and polyvinylalcohol (Table 1). Briefly, xanthan enrichments were to obtain strains that can degrade xanthan to be used as a carbon source. High temperature xanthan degrading enzymes are interesting for oil fracking, and deep sea oil applications. Polyvinyl alcohol is used as a desizing agent (thread lubricant) in textile mills and degrading this could mean less hot water usage. Strains enriched on lignocellulose would be expected to have enzymes degrading cellulose or lignin components and would be of interest in the pulp and paper industry in the case of lignin removal for paper, or the biomass based ethanol industry (cellulases and lignin degrading enzymes).

Enrichment conditions were either aerobic or anaerobic, without electron acceptor, or with ferric iron, sulfur, sulfate, arsenate as the electron acceptors, and the temperature/pH was adjusted to be close to those found in sampling sites. The environmental conditions (temperature and pH) are provided in Table 1. The media basic salt concentration were linked with the origin of the sample: thalassic or athalassic (marine or terrestrial), moreover the T, pH and Eh of the media for enrichment and pure cultures were close to the parameters of respective hot springs. The combination of different environmental samples, different substrates and different incubation conditions

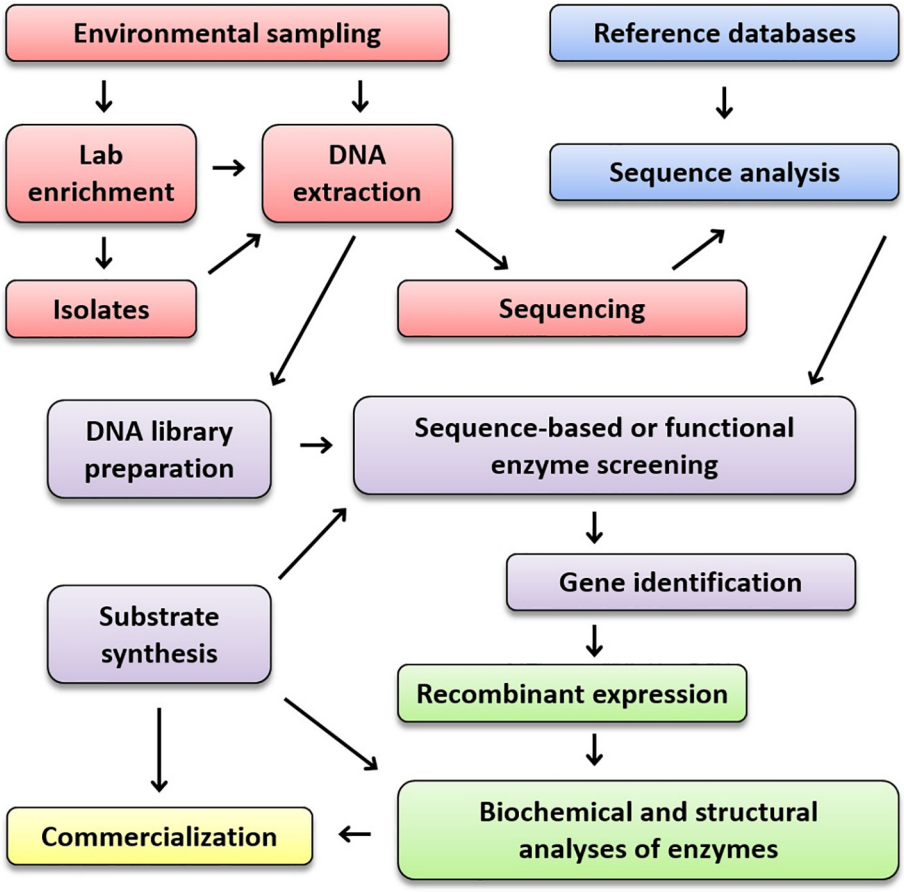


Fig. 1. Schematic workflow with exploration of the biodiversity of hot environments using different mining approaches (red boxes), involving the development of novel bioinformatic tools and platforms (blue boxes) and different screening methodologies (light velvet boxes). In the end, enzymes were subjected to more detailed functional and structural characterization (green boxes) and highly promising enzymes were patented for future potential commercialisation (yellow box). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

resulted in a huge number of enrichment cultures. Detailed recipes of culture media utilized for enrichment have been described (Zarafeta et al., 2016b; Gavrilov et al., 2016; Kovaleva et al., 2015; Podosokorskaya et al., 2014). Although more than one thousand enrichments were set up, < 10% of these survived three transfers. Only these, so-called, stable enrichment cultures, have been subsequently used for DNA isolation, for sequencing 16S rRNA genes and for isolating pure strains. Table 1 shows an overview of these strains, isolated from these stable enrichment cultures. Whereas most of the isolates show high 16S rRNA sequence identities to known species, several appeared to be quite novel and a few were published as novel taxa (Podosokorskaya et al., 2014. Kovaleva et al., 2015).

3. Sequencing and bioinformatics

3.1. Metagenomic sequencing

In early environmental DNA sequencing, Sanger sequencing was the methodology of choice, similar to sequencing individual genomes. While high quality sequence reads could be obtained, the major disadvantages were high costs, labour intensive library preparations and the lack of sufficient sequencing depth for finding rare sequences in the metagenomes.

Today, the most widely used sequencing technology for metagenomics and microbial genomics/transcriptomics is short read sequencing, for example using sequencers from Illumina, which offer a wide range of instruments with various options for sequencing depth and read length with a current maximum of 2 × 300 nt paired-end read length. Another option are the third-generation sequencers from Pacific Biosciences or Oxford Nanopore, which offer longer reads, but have a lower throughput and higher error rates (Goodwin et al., 2016). Direct sequencing of metagenomes, especially from habitats with high diversity, requires sufficiently high throughput in order to capture also low-abundant organisms in the sample. Sequencing data for HotZyme were generated either by the Roche/454 platform or by Illumina sequencing. In the latter case, 2 × 100 bp paired-end Illumina libraries were used for metagenomes and 2 × 90 bp were used for isolate genomes, and sequencing was done at BGI/Shenzhen. Detailed descriptions of the HotZyme metagenomes, including environmental conditions, sampling methods and depth, as well as biodiversity, can be found in two publications (Menzel et al., 2015; Gudbergsdottir et al., 2016).

3.2. Assembly

Sequence assembly refers to the reconstruction of the original chromosomes and plasmids from the short read data. In metagenomics, assembly is typically performed *de novo*, which means that no additional information from existing reference genomes is used for the assembly. For the assembly of millions of short sequencing reads, assembly programs, for example Velvet (Zerbino and Birney, 2008) or Spades (Bankevich et al., 2012), use de Bruijn graphs, which connect consecutive *k*-mers from the reads (Pevzner et al., 2001). This graph is then traversed for generating long contiguous sequences ("contigs"). The more complex assembly of metagenomes compared to a single individual genome is due to the fact that sequencing reads are derived from genomic fragments of multiple genomes of varying abundances. Therefore, assemblers need to avoid creating chimeric contigs that are composed of sequencing reads originating from different species, e.g. the de Bruijn graph is decomposed by the MetaVelvet assembler (Namiki et al., 2012) into smaller subgraphs of similar coverage which represent the constituent genomes. Recent metagenomic assemblers using de Bruijn graphs are Meta-Spades (Nurk et al., 2017) and Megahit (Li et al., 2015), which is particularly suited for assembling complex metagenomes of large size due to its reduced memory requirement. The

Fig. 2. Selected sample sites for mining natural thermal environments: A) Kunashir in situ, Dr. E. Taranov is gratefully acknowledged for providing Fig. 2A, B–D) Kunashir, E) Kamchatka Sun spring (Solnechniy), Russia, F—H) Yellowstone National Park, USA: (F) high temperature sulfur chloride hot spring, (G) acidic hot spring and (H) sulfuric hot spring.

length of the *k*-mers determines the number and length of assembled contigs, whose wide variation depends on what is chosen for *k*. For Illumina-sequenced metagenomes several assemblies have been generated with MetaVelvet using a range of *k*-mers and then the contigs and unassembled reads from each assembly have been merged into one final assembly per metagenome (Menzel et al., 2015).

Another approach for prevention of chimeric contigs and better recovery of less abundant genomes is to first split the sequencing reads into distinct bins that denote the individual taxa/genomes and then assemble the reads within each bin separately. This so-called binning can be based on various measures, such as the *k*-mer coverage of the sequencing reads as, for example, implemented in the LSA program (Cleary et al., 2015).

An option for increasing the recovery of protein sequences from the metagenomes is the direct assembly of protein sequences by translating sequencing reads individually followed by finding overlaps between the resulting amino acid sequences, for example with the SPA program (Yang et al., 2015). This approach can be beneficial for genomes of low abundance in the metagenome, because the effect of point mutations and sequencing errors is lower on protein-level compared with nucleotide-level so that assembly on protein-level can compensate low coverage.

3.3. Gene prediction

After the assembly of reads into contigs, *ab initio* gene prediction programs are used to find open reading frames (ORFs), which are likely to be genes coding for proteins. These programs typically employ machine learning algorithms, which are trained on databases of known genes, for example MetaGeneMark (Zhu et al., 2010) or MetaGeneAnnotator (Noguchi et al., 2008). A third commonly used program, Prodigal (Hyatt et al., 2010), does not use pre-computed gene models, but finds ORFs from the sequences between predicted start/stop codons. Since there are differences in the methodologies and the training

Table 1
Isolated pure bacterial and archaeal strains growing on polymeric substrates.

Name	Origin	Substrate, (concentration g/l)	T (°C)	pH	Closest relative at the time of isolation and 16S rDNA identity (b. – Bacteria, a. – Archaea)
7T	Iceland	Xylan, (10)	55	7.0	b. *Thermoanaerobacterium aciditolerans* 99%
7Tnr.1	Iceland	Xanthan, (0.5)	55	7.0	b. *Thermomicrobium roseum* 91%
7Tnr1 A	Iceland	Xanthan, (0.5)	55	7.0	b. *Geobacillus vulcani*
7Tnr.2	Iceland	Xanthan, (0.5)	55	7.0	b. *Cohnella laevinbosi/thermotolerans* 97%
6Tnr.2	Iceland	Xanthan, (0.5)	85	7	b. *Thermus aquaticus* 96%
6Tnr.3	Iceland	Xanthan, (0.5)	85	7	b. *Thermus aquaticus/thermophilus* 95%
8Tnr.3	Iceland	Xanthan, (0.5)	85	7.0	b. *Thermus antranikianii* 99%
2T 5 2 and Is2-7*6.2	Iceland	Xanthan, (0.5)	55	7.0	b. *Meiothermus silvanus* 99%
7T nr2,1 7T nr4,1 7T nr4,2	Iceland	Starch, (5)	55	7.0	b. *Meiothermus silvanus* 98%
2319x	Russia, Kunashir	Xylan, (10)	85	6.0	a. *Thermococcus alcaliphilus/aegaeus* 99%
2319cl	Russia, Kunashir	Sodium carboxy-methyl-cellulose, (2)	75	7.0	b. *Caldicellulosiruptor owensenis/hydrothermalis* 98%
8-7 nr.1	China	Xylan, (10)	78	7.0	b. *Dictyoglomus* sp. 99%
8-7 nr.2	China	Xylan, (10)	78	7.0	b. *Fervidobacterium islandicum* 98%
2410	Russia, Kamchatka	Polyvinyl-Alcohol, (5)	80	6.0	a. *Sulfolobus islandicus* 99%
DG#1 3,2	Denmark	Xanthan, (0.5)	55	7.0	b. *Paenibacillus ginsengihumi* 92%
DG#1 4,1	Denmark	Xanthan, (0.5)	55	7.0	b. *Cohnella laeviribosi* 97%
Is3-14,2	Iceland	Xanthan, (0.5)	55	7.0	b. *Thermus igniterrae* 98%
Is3-24,1 Is3-24,6 Is3-24,4	Iceland	Xanthan, (0.5)	55	5.0	b. *Alicyclobacillus sendaiensis* 98%
Is3-23,3 Is3-23,4	Iceland	Xanthan, (0.5)	70	5.0	b. *Alicyclobacillus acidocaldarius* 99%
DG#1 1,1	Denmark	Xylan, (10)	55	7.0	b. *Brevibacillus thermoruber* 99%
DG#1 2,1 DG#1 2,2	Denmark	Xylan, (10)	70	7.0	b. *Geobacillus vulcani* 98%
DG#1 3,2	Denmark	Xylan, (10)	55	7.0	b. *Brevibacillus thermoruber* 98%
DG#1 3,2	Denmark	Xylan, (10)	70	7.0	b. *Geobacillus thermoglucosidasius* 99%
Is3-23,2	Iceland	Xylan, (10)	70	5.0	b. *Alicyclobacillus sendaiensis* 98%
Is3-23,4	Iceland	Xylan, (10)	70	5.0	b. *Alicyclobacillus sendaiensis* 99%
Is3-24,4 Is3-24,7	Iceland	Xylan, (10)	55	5.0	b. *Alicyclobacillus sendaiensis* 99%
Is3-24,5	Iceland	Xylan, (10)	55	5.0	b. *Alicyclobacillus sendaiensis* 98%
Is3-21,2	Iceland	Xylan, (10)	55	7.0	b. *Brevibacillus thermoruber* 99%
Is3-21,4.1	Iceland	Xylan, (10)	55	7.0	b. *Geobacillus* sp. 99%
Is3-21,4.3	Iceland	Xylan, (10)	55	7.0	b. *Geobacillus thermoleovorans* 99%
Is3-23,1	Iceland	Xylan, (10)	70	5.0	b. *Geobacillus thermoleovorans* 99%
Is2-8	Iceland	Xylan, (10)	70	6.0	b. *Geobacillus kaustophilus* 97%
Is3-21,3	Iceland	Polyvinyl-alcohol, (0.5)	55	7.0	b. *Thermus brockianus* 98%
Is2-7*	Iceland	Polyvinyl-alcohol, (0.5)	55	6.0	b. *Thermus brockianus* 99%
It-5	Italy	Polyvinyl-alcohol, (0.5)	78	7.0	a. *Staphylothermus hellenicus* 96%
It-5.1	Iceland	Gelrite, (5)	78	5.0	a. *Sulfolobus shibataea* 99%
7T	Iceland	Polyethylene terephthalate, (2)	70	7.0	b. *Rhizobium leguminosarum* 91%
BPi-2ag	Russia, Baikal Rift	Agarose, (5)	54	7.3	b. *Caloramator australicus* 97%
BPi-4-40	Russia, Baikal Rift	Xylan, (2)	40	7.3	b. *Paenibacillus lautus* 99%
2842	Russia, Baikal Rift	Xanthan, (0.5)	47	7.5	b. *Phycisphaera mikurensis* 80%
2918	Russia, Kamchatka	Xanthan, (0.5)	54	6.0	b. *Phycisphaera mikurensis* 80%
Rift-s3	Guaymas Basin	Microcrystalline cellulose (10)	65	6.5	b. *Thermosipho atlanticus* 96%

sets of each program, the sensitivity of ORF annotation could be improved by combining the output of several such programs and is for example done in the IMG/M annotation pipeline (Markowitz et al., 2012). The use of multiple programs could be especially beneficial for sequencing of unknown extremophiles, whose gene properties (like codon usage and GC-content) likely deviate from the known genes that are used for training the prediction algorithms. In this case, one would preferably have too many falsely predicted ORFs, which will later remain unannotated, rather than missing some ORFs that might yield interesting proteins. *Ab initio*gene prediction can also be performed on raw sequencing reads, without assembly, where the predictors also need to compensate for sequence truncation and sequencing errors. In that case, the program FragGeneScan showed a higher sensitivity than the above mentioned programs (Trimble et al., 2012) and the precision can be improved by combining several predictors (Yok and Rosen, 2011). However, this will usually only yield truncated protein sequences, due to the short read length when using Illumina or 454 sequencing.

3.4. Functional annotation

After prediction of putative ORFs, the protein sequences need to be compared to known proteins or genes in order to learn about their putative functions. Today, several different protein databases can be either downloaded or used on-line, for example UniProt, IMG, and the NCBI protein database NR. Additionally, several annotation pipelines are available either for online use, for example RAST and IMG/M, or for a local installation, such as Prokka (Seemann, 2014), which rely on different reference databases and sequence comparison programs. A major hurdle for the annotation of genes from extremophiles is the lack of available reference protein sequences, since most known microbial genes originate from culture-dependent experiments, whereas many extremophiles are still uncultivated.

Even if it is possible to find an alignment to a known protein

sequence, it is still likely that the known protein has no annotated function (often called “hypothetical protein”). However, it is possible to annotate ORFs with no easily detectable homologue by using more sensitive search strategies (Lobb et al., 2015) or extending the search to more reference databases, e.g. Interpro, or to databases focused on certain classes of proteins. For example, the CAZY database contains enzymes from carbohydrate pathways, whereas the MEGARes (Lakin et al., 2017) and CARD (Jia et al., 2017) databases contain proteins related to antibiotic resistance.

Another option for functional annotation is to identify specific domains, e.g. DNA binding domains, by the use of the database Pfam (Finn et al., 2014), or by looking for specific sequence features, such as transmembrane helices or secretion signals (Käll et al., 2007). A protocol for these custom search strategies is outlined in Toshchakov et al. (2017).

Programs for the comparison of protein sequences include NCBI BLAST+ (Camacho et al., 2009), DIAMOND (Buchfink et al., 2014), and MMSeqs (Hauser et al., 2016), which differ in speed and sensitivity. For the most sensitive search, profile comparison programs, such as Psi-BLAST (Altschul et al., 1997), HHPred (Söding et al., 2005), or HMMER (Eddy, 2011), can possibly find more distant homologues in the protein databases.

The output of the various annotation programs for all genes identified from our sequenced metagenomic and culture-based samples was combined and presented in the web-based portal ANASTASIA, based on the popular Galaxy platform (Afgan et al., 2016), thus facilitating easy access and search (Theodoros Koutsandreas, Efthymios Ladoukakis, Eleftherios Pilalis, Dimitra Zarafeta, Peter Menzel, Anders Krogh, Georgios Skretas, Fragiskos Kolisis, Aristotelis Chatziioannou, manuscript in preparation). The ANASTASIA portal is open to the public upon registration for academic purposes through the following links: http://motherbox.chemeng.ntua.gr/anastasia_dev/ & https://bio.tools/ANASTASIA.

3.5. Taxonomic classification

The community composition for some of the sequenced hot spring metagenomes is described quantitatively by Menzel et al. (2015). To this end, we assigned taxa to assembled contigs by using Rapsearch2 (Zhao et al., 2012), MEGAN (Huson et al., 2011) and re-mapped reads to the assembled contigs to calculate taxon abundances. Additionally, the samples were searched for 16S rRNA genes using the ARB-Silva SSU database (Quast et al., 2013). Taxonomy assignment from isolated genomes was performed by NCBI BLAST with phylogenetic markers found in these genomes (as 16S and 23S rRNA genes, ribosomal proteins etc.) as queries using the “type material” option. During the project, we also developed a new fast and sensitive method, “Kaiju”, for directly quantifying taxon abundances from metagenomic sequencing reads (Menzel et al., 2016). The main advantage of Kaiju is an increased sensitivity of the taxonomic classification by using sequence comparison on the protein-level instead of comparing nucleotide sequences. This is important for metagenomes from environmental samples, especially from extreme environments, because there are only few reference genomes available for extremophiles and therefore the classification method has to overcome potentially high evolutionary distances when comparing sequences. Further, the method is as fast as the fastest state of the art programs, allowing the classification of an entire metagenome within minutes.

4. Gene expression and screening

4.1. Recombinant libraries for functional screening

In order to screen for genes that encode enzymes in a recombinant system, one must first create a DNA library (Fig. 1) suitable for such screening in a suitable host. To enhance the probability of identifying relevant genes in such a recombinant library screening, a number of factors must be considered. Choosing the appropriate expression system paired with the appropriate screening host is essential. In the HotZyme project, plasmid libraries were produced in vectors pUC18 and pRham (Lucigen) designed for screening in *Escherichia coli*. The main reason for choosing an *E. coli* based system was that it matched the criteria for the HotZyme project:

1. The screened DNA came from environmental samples, enriched samples and single organism isolates. Especially for the environmental and enriched samples, a recombinant host with very high transformation rates is desirable as the hit rate for a particular enzyme gene falls as the relevant genome carrying the gene is diluted with genomes that do not encode such an enzyme.
2. *E. coli* has been shown to be highly accepting of foreign Shine-Dalgarno translation initiation sites. Enzymatic activities have been screened for and recovered from Gram positive bacteria, Archaea, and even plant cDNA libraries. In the case of cDNA libraries, it was shown that translation initiation often occurred using the 5′ untranslated region of the cDNA (Schnorr et al., 1994; Hoff et al., 1995; Schnorr et al., 1996a; Schnorr et al., 1996b; Vincenzetti et al., 1999; Becker et al., 2004).
3. The methylation status of donor DNA can be addressed by the use of *E. coli* strains developed for receiving foreign DNA. Such strains lack the restriction systems McrA, McrBC and Mrr found in wild type *E. coli* and therefore can accept foreign DNA without destroying it (www.neb.com).
4. In addition to considerations of library quality, average insert size, library complexity and percentage of recombinants carrying inserts, an additional quality parameter was explored in the HotZyme project. Purine biosynthesis, with very few exceptions, is required of all living organisms and requires 14 enzymatic steps (Zalkin and Dixon, 1992). The single reactions steps of the enzymatic pathway are identical and stable *E. coli* auxotrophs lacking specific purine biosynthetic enzymes have been used as a basis for screening such enzymes from foreign sources (Schnorr et al., 1994). *E. coli* strain SØ3274 lacks the PurD (GAR synthetase) enzyme and was used to determine the “screening quality” of libraries produced in the HotZyme project by transformation of the strain SØ3274 with a recombinant library and plating on minimal media without purine.

To prepare recombinant plasmid-based expression libraries of HotZyme enrichments and isolates, genomic DNA from different sources was cloned into pRham utilizing the Shotgun Expression Cloning kit (Lucigen). For cloning into the pUC18 vector, DNA was subjected to partial digestion with *Bsp*143I and *Hin*1II and fragments of > 2 kb were gel extracted and ligated with *Bam*HI/*Sph*I digested pUC18. The ligated DNA was transformed into the highly competent *E. cloni*® 10G SUPREME cells (Lucigen) (Zarafeta et al., 2016b).

4.2. Functional in vitro screening of expression libraries

For successful functional screening of *E. coli* expression libraries for hydrolytic enzymes, several factors need consideration:

1. A suitable assay is needed to detect the desired enzyme activity, for example by use of a chromogenic (surrogate) substrate.
2. The substrate must be accessible to the expressed enzyme, either because the substrate is taken up by the cell, the enzyme is secreted or it is released by cell lysis, although leakage of highly expressed intracellular enzymes into the extracellular space may occur.
3. To achieve a good signal-to-noise ratio, the screening host should not contain significant endogenous activity towards the screening substrate.

For typical expression library sizes (10^4–10^5 clones), a commonly

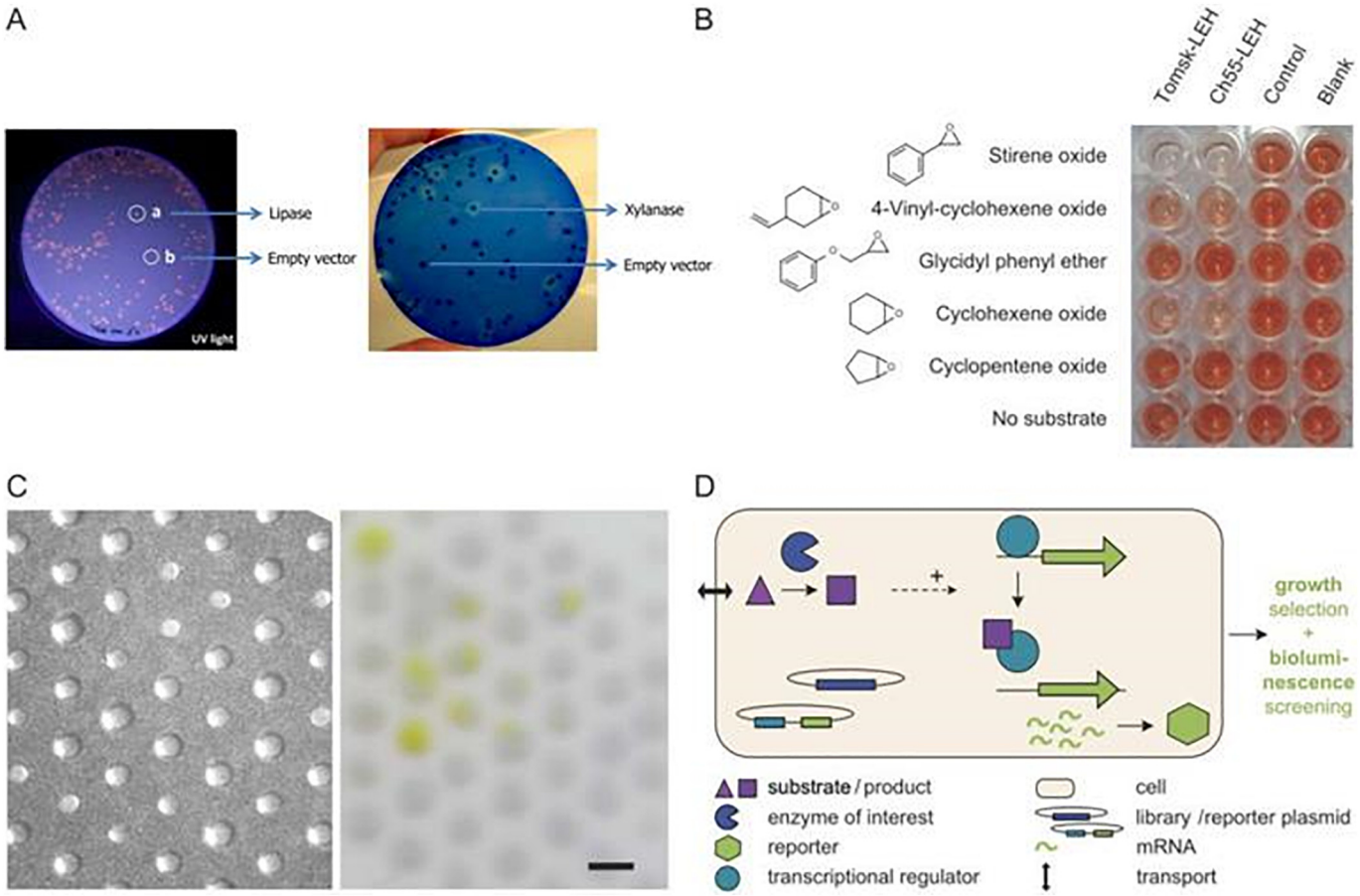


Fig. 3. Overview of the different functional screening strategies. (A) Agar plate based screening. (B) Microtiterplate-based screening. Detection and substrate profiling of epoxide hydrolase activity with the colorimetric adrenaline assay (Cedrone et al., 2005). (C) Microbial culture chip (MCC) based screening. Left: light microscope image of a small section of an MCC with 100 µm circular compartments, inoculated with a mixture of *E. coli* expressing an esterase or carrying an empty vector. Right: a similar area is shown after transfer of the microcolonies onto nitrocellulose and incubation on a filter paper soaked with assay buffer containing substrate and the pH indicator bromophenol blue. The scale bar is approximately 300 µm. (D) Whole-cell bioreporter-based screening. Adapted from van Rossum et al. (2013). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

used and robust approach is to examine bacterial colonies on agar plates that contain a screening substrate (Zarafeta et al., 2016b; Popovic et al., 2017). The substrate is often a colourless compound that produces a colour reaction, either because a chromophore or fluorophore is released upon enzymatic hydrolysis, or because the enzyme product can be indirectly detected by means of a colour reaction. The activity of endoglucanases such as cellulase or xylanase can be detected by the use of polysaccharides conjugated to remazol brillant blue R (RBB) or cross-linked with azurine (Fig. 3A). Alternatively, substrate hydrolysis in plates containing unlabelled beta-D-glucans (such as carboxymethylcellulose) or D-xylan can be detected with the dye congo red (Teather and Wood, 1982) that strongly interacts with the polysaccharides. Numerous substrates are available for screening esterases and lipases. While the specific choice may depend on the desired specificity, it also strongly affects the success rate of the screen (Ferrer et al., 2016). As an example, esterases and lipases can be readily screened for on agar plates that contain emulsified triglycerides, tributyrin for esterase and triolein or olive oil for lipase activity. A clear halo can be observed in the turbid tributyrin medium around the esterase producing colonies, exemplified by the esterase EstDZ1 from a new *Dictyoglomus* isolate (Zarafeta et al., 2016b). Lipase activity screening can be aided by the addition of rhodamine B, a fluorescent dye that is presumed to form a complex with free fatty acids yielding bright orange fluorescence (Kouker and Jaeger, 1987). Moreover, several classes of hydrolases produce an acid or base as reaction product (such as lactonases, esterases and lipases) and the concomitant pH change can be easily measured indirectly with pH indicators. Consequently, such generalized assays have been very popular and versatile, as screening for substrate scope can be performed with the real substrates and does not require additional synthetic steps for their labeling to chromogenic or fluorogenic enzyme substrates. If carefully designed, the assay can quantitatively measure an enzyme's reaction rate. Similarly, several enzymes including lactonases and esterases identified by the HotZyme consortium were characterized with a pH indicator assay using newly synthesized racemic and enantiopure substrates (see Section 5, Functional and structural enzyme characterization). Direct screening on agar plates may not be possible if the cells need to be lysed to release intracellular enzymes, the substrate is toxic to the screening host or if a multistep assay is required to detect activity. An example for the latter is the detection of epoxide hydrolase activity with a colorimetric adrenaline-coupled assay (Cedrone et al., 2005). In such cases, screening can be conducted in microtiter plates. Individual library clones are picked and stored in master plates and transferred to new plates for growth and activity measurements (Fig. 3B). While screening on agar plates has the above mentioned limitations, microtiter plate based screening at high throughput can be resource demanding due to the need of robotics for a large number of plates. Several alternative solid phase approaches have been presented in the literature to tackle these challenges. One possibility is replica plating bacterial colonies onto a nitrocellulose membrane and to perform cell lysis (rapid freezing, chloroform vapour) and enzyme assay on the membrane. Activity in the transferred colonies is detected by placing the membrane

on filter paper soaked with a suitable substrate. A miniaturized version of this method using microbial culture chips (MCC, Ingham et al., 2007; Caton et al., 2017) has been developed. MCCs are thin ceramic (porous aluminium oxide) membranes subdivided into microwells by a patterned polymer layer, which is laminated onto the membrane, e.g. a MCC having compartments with 100 µm diameter and spacings of 200 µm results in a colony density of > 1000 microcolonies/cm^2. An *E. coli* expression library inoculated on the MCC grow as microcolonies and can be replica plated onto nitrocellulose membranes for screening on a filter paper soaked with the screening substrate of interest. Detection is done visually using a low magnification light or fluorescence microscope (Fig. 3C). Other microcolony-based approaches have been developed that require dedicated instruments or robotics (Youvan et al., 2002), but offer solid phase screening paired with signal quantification. A further increase in throughput (> 10^7 clones) can be achieved by the combination of in vitro compartmentalization (IVC) and fluorescence-activated cell sorting (FACS). In IVC, cells of a library or a system for in vitro transcription-translation are encapsulated with a fluorogenic enzyme substrate in droplets consisting of a water/oil/water emulsion. Substrate hydrolysis leads to an increase of fluorescence in the compartment and is detected in a FACS instrument that can also sort positive clones for subsequent recovery (Yang and Withers, 2009). Originally developed for directed evolution studies, such methods are translatable to the screening of (meta)genomic expression libraries. In a similar approach, microfluidic screening platforms are used to sort droplets that are fluorescent and therefore contain active enzyme (Baret et al., 2009), for example specific hydrolases in a metagenomic library (Colin et al., 2015). Fluorescence-based cell sorting techniques can also be used to screen libraries containing reporter genes (usually green fluorescent protein) linked to enzyme activity (see below).

4.3. Functional screening with whole-cell bioreporters

In enzyme screening, general methods for detecting enzyme activity in whole cells are of much interest. A living microorganism which is able to provide a detectable phenotype by switching on a reporter, when the enzymatic reaction product is bound to a sensor molecule of this microorganism, is defined as a whole-cell bioreporter, in short: bioreporter (Hynninen et al., 2010). In contrast to screens that rely on the detectability of the product itself, this indirect measuring makes the method independent of the reaction type and allows a generalized screening of a broad range of enzymes (van Rossum et al., 2013). This also allows choosing the type of reporter based on the desired detection method which enables high-throughput screening. This is of great interest for coping with large metagenomic or enzyme variant libraries in screening approaches (van Rossum et al., 2013; Zhang et al., 2015). The reporter applied most often is GFP (Zhang et al., 2015), because it is simple, no substrates are needed and, most importantly, it can be coupled to high-throughput FACS. Another high-throughput approach is to use a reporter that enables cell growth through for example antibiotic resistance or auxotrophy complementation (Zhang et al., 2015). By this approach only positive cells remain in the variant pool. Moreover, in contrast to FACS, it does not require expensive equipment. Although all cells that survive are potentially interesting, a high false positive rate is a risk (Raman et al., 2014; Rogers and Church, 2016). To limit this rate one could tune the selection system, combine negative and positive selection or combine a reporter for selection and a reporter for screening (Raman et al., 2014; Mahr and Frunzke, 2016; van Rossum et al., 2017). Other commonly used reporters are LuxCDABE, LacZ and MtrB.

The sensor part of the bioreporter couples the enzymatic reaction to the reporter and can consist of various kinds of biomolecules, either protein or RNA (riboswitches). A wide variety of protein-based systems have been described of which sensor parts include enzymes, transcriptional regulators, extracytoplasmic function (ECF) sigma factors, two-component systems, Periplasmic Binding Proteins (PBPs), or fluorophore-containing proteins (e.g. proteins that allow for Fluorescence Resonance Energy Transfer (FRET)). Design and optimization of new sensors is, however, challenging and time-consuming. Depending on the reporter, transcriptional-regulator- or riboswitch-based screens are high-throughput and their sensor design is, currently, a bit easier compared to that of other types. The transcriptional regulator or the riboswitch binds the product of the enzymatic reaction and undergoes a conformational change, allowing transcription or translation of the reporter, respectively. Riboswitch-based sensors have a modular design, their binding and folding properties are understood in depth, aptamers can be de novo generated in vitro and they provide less burden to the cell, because no transcription is required. However, translation to in vivo is not always straightforward. Transcriptional-regulator-based sensors on the other hand, have a higher output fold change upon addition of the small molecule and can be applied for a more diverse range of small molecules, because of their more diverse chemistry, consisting of amino acids versus nucleotides. More information about bioreporters can be found in the reviews of Michener et al. (2012), van Rossum et al. (2013), Boock et al. (2015), Eggeling et al. (2015), Zhang et al. (2015), Liu et al. (2015) and Mahr and Frunzke (2016).

At the moment, transcriptional-regulator-based bioreporters are the most successful (Fig. 3D). In the majority of these cases, the transcriptional regulators are natural regulators, often originating from the model organism *E. coli* or from soil bacteria like *Pseudomonas putida*. However, not for all enzymatic products a natural transcriptional regulator is available that has the desired characteristics needed for screening. Therefore, some bioreporters are based on transcriptional regulators that have been engineered via directed evolution (Reed et al., 2012; Tang et al., 2013) or via computational design (de los Santos et al., 2016; Jha et al., 2016). Transcriptional-regulator-based bioreporters can also be used for finding novel hydrolases. Uchiyama and Miyazaki (2010) found three novel benzamide converting amidases by co-cultivating metagenomic library cells (created with DNA from activated sludge) and sensor cells that contain the benzoate-responsive transcriptional regulator BenR as sensor and GFP as reporter. For a library size of 96,000 clones, the 96-well format was reasonable, but for larger libraries the combination with FACS is needed. Several studies use FACS in combination with the phenol-responsive transcriptional regulator DmpR to identify novel phenol releasing enzymes from metagenomic libraries (created with DNA from sea tidal flat sediments) or enzyme variant libraries. These enzymes include several hydrolases like phosphatases, lipases, cellulases and methyl parathion hydrolases (Jeong et al., 2012; Choi et al., 2014; Kim et al., 2016). Jha et al. (2016) engineered the 4-hydroxybenzoate-responsive transcriptional regulator PobR to respond to *p*-nitrophenol. This PobR variant allowed discrimination between cells with different levels of phosphotriesterase activity on paraoxon. All these examples concern hydrolases from mesophilic environments. To extend the application of bioreporters to thermostable hydrolases, two approaches are imaginable. A thermophilic host could be used as bioreporter, since more and more thermophilic organisms become available for genetic engineering and are also used for enzyme screening, such as *Thermus thermophilus* and *Geobacillus thermodenitrificans* (Leis et al., 2015; Daas et al., 2018). However, a simpler approach is the use of the currently applied mesophilic hosts, as long as expression is not an issue. The activity of thermostable enzymes in mesophilic hosts can be sufficient for functional screening (Tirawongsaroj et al., 2008).

The HotZyme project provides to our knowledge the only case of a bioreporter used to detect a thermostable enzyme (van Rossum et al., 2017). As proof of principle for using this bioreporter for the detection of a thermostable enzyme, the readily available L-arabinose isomerase of *G. thermodenitrificans* was chosen, because the initially developed bioreporter was L-arabinose specific. L-arabinose isomerase-expressing cells were enriched more than a millionfold, when the substrate L-ribulose was present, using both selection and screening. The L-

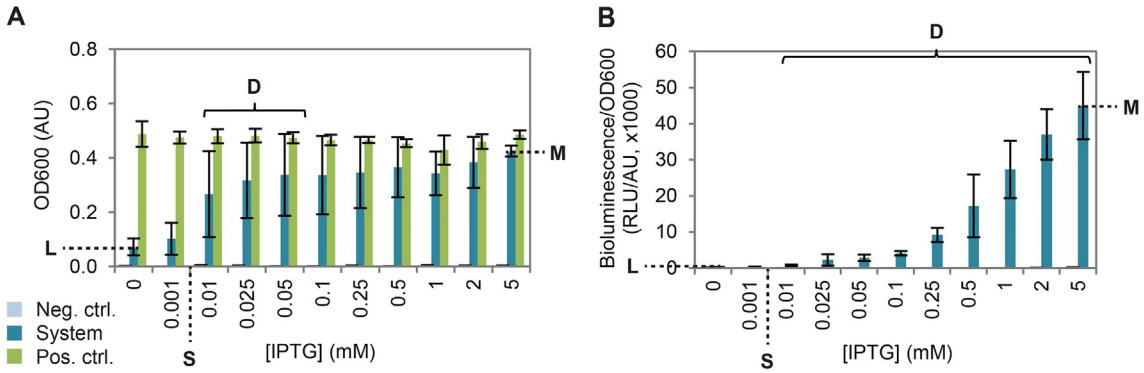


Fig. 4. LacI-based bioreporter with double reporters. (A) Selection based on leucine auxotrophy complementation. The plasmid-encoded reporter gene *leuB* was induced by various concentrations of the inducer IPTG. Bacteria were grown in M9 medium for 48 h. (B) Screening based on bioluminescence. The plasmid-encoded reporter operon *luxCDABE* was induced by various concentrations of the inducer IPTG. Bacteria were grown in LB medium for 4.5 h. System: auxotroph *E. coli* BW25113 Δ*lacI* Δ*leuB* Δ*recA* (LLR) with the regulator-reporter plasmid. Neg. ctrl.: auxotroph LLR with the regulator-reporter plasmid with a frameshift in *leuB* (A) or *luxA* (B). Pos. ctrl. (only in (A)): non-auxotroph *E. coli* BW25113 Δ*lacI* Δ*recA* (LR) with the regulator-reporter plasmid with a frameshift in *leuB*. The data are an average of three independent experiments (standard deviation indicated). L, leakiness; M, maximal signal; D, dynamic range; S, sensitivity. For L and M, the neg. ctrl. signal was subtracted from the system signal.

arabinose-responsive transcriptional regulator AraC activates transcription of the kanamycin resistance marker (*kan*) and the luciferase genes (*luxCDABE*), allowing to reduce rapidly the metagenomic or mutant library size in one step, based on growth in the presence of kanamycin, and to exclude subsequently false positives, based on bioluminescence. This bioreporter with dual reporters therefore tackles the large false positive rate often encountered for growth-based selection and makes quantification of positive variants possible. To show adaptability of the system to other transcriptional regulators and thus to other enzymatic products, two approaches were taken. One approach focused on engineering the ligand specificity of AraC towards other monosaccharides, e.g. D-xylose. To this end, *araC* variant libraries were made by combinatorial site-saturation mutagenesis and variants were selected using the dual reporter system (van Rossum et al., unpublished results). In the second approach, AraC was replaced by allolactose- or IPTG-responsive LacI (van Rossum et al., unpublished results). Different system versions were developed that varied in plasmid copy number (low p15A or medium ColE1) and selection reporter (*leuB* or *kan*). The low copy system with LeuB, allowing selection based on leucine auxotrophy complementation, was the best system (Fig. 4). It showed a good sensitivity (0.001–0.01 mM) and a low leakiness in selection, which is very important to reduce the false positive rate, while not missing any positives. In screening, it had a good fold change of the maximal signal over the leakiness (three orders of magnitude), a satisfactory dynamic range (two orders of magnitude) and a good sensitivity (0.001–0.01 mM), all important for verification and quantification. These characteristics are comparable to values of other transcriptional-regulator-based bioreporters (van Rossum et al., 2017). As predicted upfront (van Rossum et al., 2017), a different regulator indeed demands for some optimization followed by characterization. For the best performance, the AraC- and LacI-based systems require other components like medium versus low copy number and KmR versus LeuB as selection reporter, respectively. In addition, characteristics of the two systems differ; the LacI-based system has a better sensitivity and a higher fold change of maximal signal over leakiness, but its dynamic range for selection is less. This study shows the adaptability of the dual bioreporter to another transcriptional regulator, broadening the bioreporter's range of target molecules.

The availability of sensors and more specifically of transcriptional regulators for the target enzymatic products is the major determinant in making bioreporters widely applicable in novel biocatalyst detection. Fortunately, more and more efforts are being made to identify and characterize new transcriptional regulators. Next to regulator engineering, one approach is to pick up promoters responsive to a target molecule from promoter libraries (Mahr et al., 2016). Helper enzymes can be used alternatively to convert the target enzymatic product to a molecule for which a transcriptional regulator is known (Rogers and Church, 2016). Other issues that could hamper the applicability of bioreporters are of more general nature, like problems with heterologous expression or uptake of substrate or product by the cell. In short, efforts like increasing the number of available sensors, expanding the host and reporter repertoires and improving general issues like heterologous expression, should make bioreporter-based detection of numerous (thermostable) enzymes possible.

5. Functional and structural enzyme characterization

A wealth of sequencing data has been obtained from the newly isolated bacterial and archaeal genomes of single organisms and from metagenomes. These sequencing data led to numerous hits for targets coding for enzyme functions from enzyme subclasses of EC 3. It is important to characterize the industrially relevant thermostable hydrolase enzymes both biochemically and structurally in order to optimise their applications for industrial biocatalysis. This has been carried out for a selection of the enzymes with highest industrial interest which have been successful in overexpression at high levels and in a soluble form using different commercially available strains of *E. coli* as hosts. This has had a high success rate using synthetic genes that are codon optimized for expression in this host. The enzymes can be rapidly purified using a His-tag and nickel affinity chromatography. Once these enzymes are required in commercial quantities other hosts including *Aspergillus* sp. will be used where the enzyme is transported to the growth media. Many other enzymes remain to be characterized due to time restrictions of the HotZyme project. Biochemical and structural characterization including both molecular modelling and crystallographic structure determination have been carried out on a selection of different hydrolase enzymes (outlined in Table 2) within the project, including thermostable carboxylesterases, enol lactonases, quorum sensing lactonases, gluconolactonases, limonene epoxide hydrolases, αβ hydrolase fold epoxide hydrolases and two novel cellulases. An essential prerequisite for the discovery of novel enzyme functions is the synthesis of the corresponding enzyme substrates and a selection of representative substrates used for the characterization of the novel hydrolases is shown in Fig. 5.

Over 100 other enzymes were identified but were not selected for further characterization within the project due to over-expression problems, were considered lower priority or were similar to enzymes already described in the literature. Since all proteins selected were from

Table 2
Enzymes selected for biochemical and structural characterization.

Enzyme	EC class	Gene Bank Accession Number	PDB code
Lactonases			
Quorum sensing Lactonase	EC 3.1.8.1	*Vulcanisaeta* sp.CP002529	4RDZ, 4RE0, 4RDY
Enol Lactonase	EC 3 1.1.24	*Carboxydothermus* sp.CP000141	Not deposited
Carboxyl esterases			
AF-Est2 Family 6	EC.3.1.1.1	*Archaeoglobus* sp. AE000782	5FRD
TtEst		KR002593	4UHC,4UHD,4UHE,4UHF
TtEst2		KT724966	5AO9,5AOA,5AOB,5AOC
EstDZ2	EC 3.1.1.1	KX301277	
EstDZ3	EC 3.1.1.1	KX557297	
Epoxide Hydrolases			
Tomsk-LEH	EC 3.3.2.8	KP76511	5AIF, 5AIG
CH55-LEH		KP76710	5AIH, 5AII
SIBE-EH	EC 3.3.2.9	KX505385	5NG7
CH65-EH		KX505386	5NFQ
Cellulases			
CelDZ1	EC 3.2.1.4	KT844947	5FIP
Glycosidase-Endoglucanase/Mannosidase	New multidomain enzyme	*Thermococcus* sp. CP012200	

thermophilic sources they were generally more thermostable than homologues already discovered and they showed some novelty in their structure where determined.

In order to further understand enzyme mechanism and substrate specificity to customise the enzyme properties for industrial applications, site-directed mutagenesis experiments have been carried out to optimise the substrate specificity for some of the enzymes. This has been driven by knowledge of the structures of the new esterase and epoxide hydrolase enzymes. The enzymes studied are all from thermophilic sources and as such are all thermostable and relatively stable to organic solvents. These properties provide robust industrial enzymes that do not require further optimisation as regards stability for industrial applications and have the advantage that they can be recycled for several rounds of catalytic processes without being denatured. This reduces the costs and economic feasibility for selecting a biocatalytic application rather than more unsustainable processes.

5.1. Lactonase enzymes

Lactonase enzymes are not commercially available and they are of great interest to industry, because mild and selective hydrolysis of lactones can provide simple access to enantiomerically pure products, such as chiral lactones from the enzymatic resolution of racemic lactones. However, they are relatively understudied and at the start of this project, there were no thermostable lactonases characterized. The ability to use a variety of important lactones as substrates and to convert them in a stereospecific way is a novel feature not previously studied in any detail. A bioinformatics search of the *Carboxythermus hydrogenoformans* genome was performed to discover potential lactonase genes. This search revealed an α/β hydrolase gene with 27% amino acid sequence identity to a known enol-lactonase from *Burkholderia xenovorans*. This gene has been cloned and over-expressed to purify and study the enzyme both biochemically and structurally. A rapid screening assay was developed based on monitoring lactonase activity using the indicator dye bromocresol purple for the detection of the product of the lactonase reaction, which causes a pH change. The activity of the cloned *C. hydrogenoformans* lactonase (*Ch*-LAC) was assayed against a selection of lactone substrates of interest. Activity was observed for δ-decalactone, δ-dodecalactone, γ-valerolactone, γ-butyrolactone and γ-caprolactone.

The *C. hydrogenoformans* lactonase was crystallized by the microbatch method using an Oryx Robot (Douglas Instruments) and commercially available crystal screens (Molecular Dimensions). The quality of crystals was optimized in the presence of Bis-Tris propane buffer and diffraction was observed to 2.6 Å resolution. A full crystallographic data set was collected with the crystal belonging to space group $p2_12_12_1$ with cell dimensions of a = 45.4 Å, b = 144.4 Å, c = 183.3 Å, α = β = γ = 90°. The structure was solved by molecular replacement using a hydrolase from *Pseudomonas aeruginosa* PA01 (PDB: 3OM8) as a model which shares 28% sequence identity to the *C. hydrogenoformans* protein.

Studying these crystals by X-ray diffraction has allowed the structure of this enzyme to be determined. It belongs to the α/β hydrolase fold family consisting of core and cap domains and the catalytic triad (Ser92, Asp209 and His237) located at their interface (Fig. 6) (Sayer et al., 2016, unpublished results).

Ligand docking into the active site of this enol lactonase enzyme was used to rationalise its substrate specificity. The lactonase was active towards γ-butyrolactone (GBL), γ-valerolactone (GVL) and γ-lactones with longer chains, and was not active towards the (3*S*,4*S*)-isomer of whiskey lactone (WGL) (Fig. 7).

To better understand the binding mode of the enzyme substrates and also the lack of activity towards WGL, the Autodock Vina program was used for docking the structures of these ligands into the enzyme's active site (Fig. 8). The results showed that the enzyme substrates γ-butyrolactone, (*R*)-γ-valerolactone and (*S*)-γ-valerolactone can bind in a configuration where one of their two oxygen atoms lies close to the active site Ser92 OG and another makes a hydrogen bond to the nitrogen of Leu26, which is part of the oxyanion hole. With WGL as substrate, the methyl group prevents the effective binding of the ligand in the catalytic position. In all of the WGL docked configurations one of its oxygen atoms is far from Ser92 OG in the active site and the oxyanion hole explaining the lack of lactonase enzyme activity towards WGL.

5.2. Vulcanisaeta mountnovskia quorum sensing type lactonase

Structurally and biochemically related to the phosphotriesterases (PTE), the quorum sensing lactonases (PLL) show lactonase activity and also a promiscuous PTE activity. Known members of the PLLs are from *Sulfolobus acidocaldarius* (SacPOX) and *S. solfataricus* (SsoPOX). These enzymes hydrolyse N-acyl-homoserine-lactones (AHL) which mediate bacterial communication. A BLAST sequence comparison using the archaeal *Sulfolobus* lactonases has identified a putative lactonase in the recently sequenced genome of *Vulcanisaeta mountnovskia* with 52% sequence identity. In order to determine if the *V. moutnovskia* had PTE or PLL activity cloning and overexpression of the predicted phosphotriesterase was done in *E.coli*, which was followed by protein purification and biochemical characterization (Kallnik et al., 2014).

The substrate specificity of the lactonase from *V. moutnovskia* was

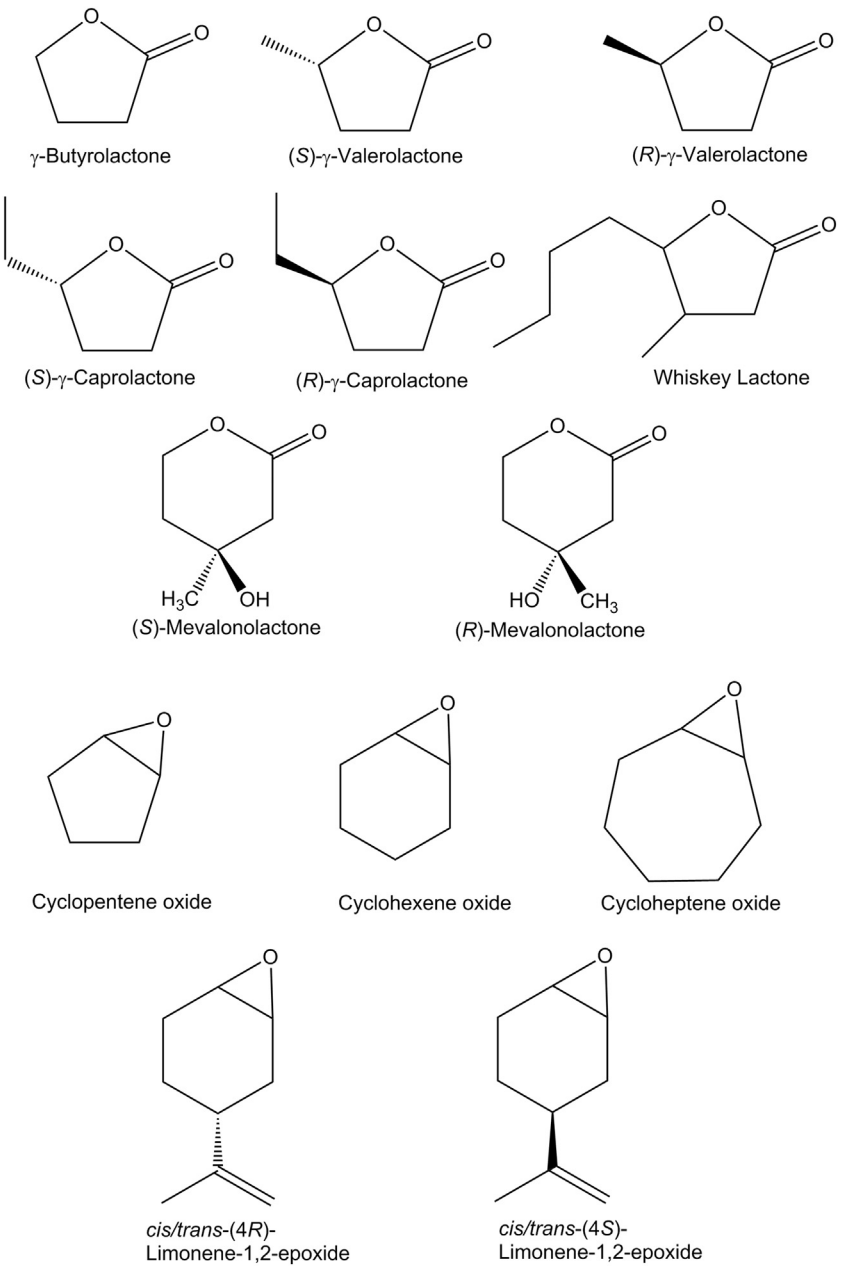


Fig. 5. Representative enzyme substrates which have been used for the discovery and characterization of novel hydrolases described in this section.

evaluated using industrially relevant substrates. The enzyme has been shown to be highly active using γ-lactone substrates with different hydrophobic substituents ranging from a methyl side chain with GVL to γ-dodecalactone. Activity was found with WGL and δ-dodecalactone and measurable activity was detected with mevalonolactone or δ-decalactone. All enantiomers of the two substrates GVL and γ-caprolactone were tested to determine stereoselectivity of the enzyme. Results indicated that while activity is seen with both isomers the enzyme seems to favour the D form of these substrates (Kallnik et al., 2014).

The *V. moutnovskia* lactonase structure (Hiblot et al., 2015) shows a substrate binding pocket lined with hydrophobic amino acid side chains and located in the enzyme structure by a bound long chain fatty acid. The homologous lactonase from *Sulfolobus islandicus* has been reported to favour γ-lactone substrates with long hydrophobic acyl chains of over four carbons in length (Hiblot et al., 2012). The substrate specificity differences between the *Vulcanisaeta* and the *Sulfolobus* enzymes have been rationalised by molecular modelling (Saneei et al., unpublished results). The substrate binding site of the *S. islandicus* lactonase revealed a number of polar residues at the entrance of the substrate pocket with hydrophobic residues lining the bottom of this pocket. This would disfavour the binding of the smaller γ-lactones in the correct catalytic position. The different distribution of hydrophobic and other residues in the active site pocket between the two related enzymes appears to be responsible their different substrate preferences.

5.3. Esterase enzymes from a novel Planctomycetes sp., Thermogutta terrifontis

Thermogutta terrifontis was found to be the first thermophilic member of the environmentally important *Planctomycetes* group (Slobodkina et al., 2015; Elcheninov et al., 2017). Cloning and overexpression of two different novel carboxyl esterases from this organism in *E. coli* have been done as part of the Hotzyme project. Biochemical and structural characterization of the first carboxyl esterase enzyme (TtEst) has shown that *p*-nitrophenyl-propionate is the best substrate among small *p*-nitrophenyl-esters and that the enzyme is stable up to

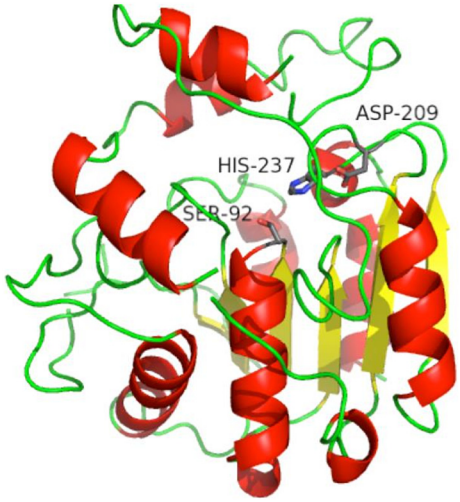


Fig. 6. The overall α/β structural fold of the *C. hydrogenoformans* lactonase enzyme with the catalytic triad characteristic of esterase enzymes shown in stick representation. Subtle differences in the aminoacids within the binding pockets determine the preference for lactone substrates.

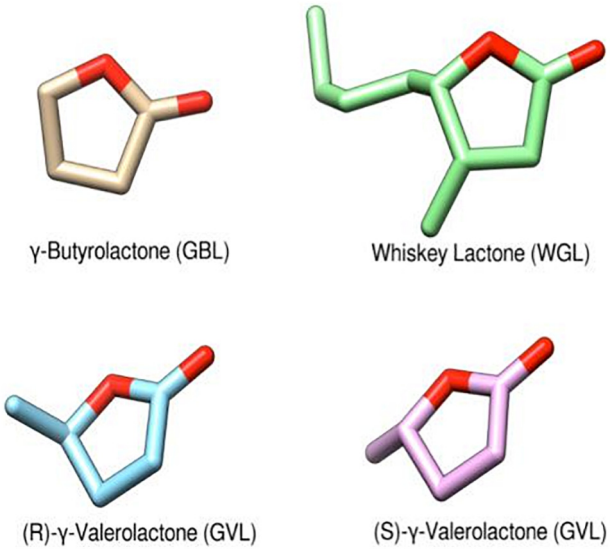


Fig. 7. The structure of the substrates used in the molecular modelling ligand docking studies of the *C. hydrogenoformans* lactonase enzyme.

high temperatures. Activity is maintained at 95% when TtEst is incubated for 60 min at 80 °C. The protein was crystallized from condition F10 of the JCSG commercial screen and crystals of the native enzyme and enzyme complexes belonged to space group P3221 (cell dimensions a = b = 43.3, c = 227.1 Å, α = β = 90°, γ = 120°). The crystals obtained diffracted to 1.1 Å resolution (Sayer et al., 2015a).

The enzyme, which is a monomer in solution, is made up of a core domain which has an α/β hydrolase fold and a cap domain which covers the active site. A catalytic triad of Ser101, His250 and Asp222 forms the active site, with the active serine located at the end of strand β5 in a tight nucleophilic elbow as seen in other esterases, and with the classic signature sequence conserved.

The TtEst enzyme structure consists of a small pocket defining what size carboxyl ester can be bound and an alcohol binding pocket, larger and closer to the active site funnel. This has been established by solving the crystal structure of the enzyme under two different conditions. The structure of TtEst obtained from crystals grown in the presence of racemic malate and imidazole, polyethyleneglycol 4000 at pH 6.0, showed additional electron density for a D-malate molecule involving its two carboxyl groups in its binding to the enzyme's active site. The proximal carboxyl group forms a hydrogen bond to the backbone oxygen of Phe35 and to the side chain of His250, while a salt bridge to Arg 139 is formed by the distal carboxyl group, which in addition forms a hydrogen bond to Tyr 105 by the phenolic oxygen. The D-malate hydroxyl oxygen, which is located close to His250 side chains and Ser101 of the catalytic triad, forms a hydrogen bond to water. Other amino acid residues located close to D-malate define the alcohol binding pocket. This feature of the enzyme is the site of *p*-nitrophenylester binding during TtEst-catalyzed hydrolysis (Fig. 9A).

When crystals were grown in ammonium acetate and polyethyleneglycol 10,000 at pH 4.0, additional electron density was observed for acetate located in the active site. The carboxyl oxygens of this acetate molecule are oriented differently, with the first one near the serine of the catalytic triad and the second one forming a hydrogen bond to the amino acid residues that makeup the oxyanion hole (Fig. 9B). The ester substrate is proposed to be in this orientation during the reaction since it locates correctly the acetate carbon. The available space in the TtEst carboxyl binding pocket limits its specificity to react with substrates containing a small carboxyl group and results in low activity with larger substrates such as *p*NP-butyrate. This study of TtEst increases understanding of the evolution of different substrate specificity for esterases. With knowledge of the crystal structure two mutant enzymes have been produced by mutation of leucine residues to smaller alanine residues. This change increases the size of the active site allowing activity towards the larger substrate *p*NP-butyrate (Sayer et al., 2015a). The L37A mutant structure, obtained by soaking the crystal in the larger butyrate substrate at pH 4.0, showed clear electron density for the larger substrate in the extended active site. The C4 atom of the butyrate is positioned equivalently to the leucine side chain of the native enzyme. The comparison of TtEst with related α/β hydrolase fold family structures showing activity towards different substrates such as lactones, γ-lactams and different esters, has revealed the subtle modifications in the active site that account for the different substrate specificities. A BLAST search of the TtEst amino acid sequence against

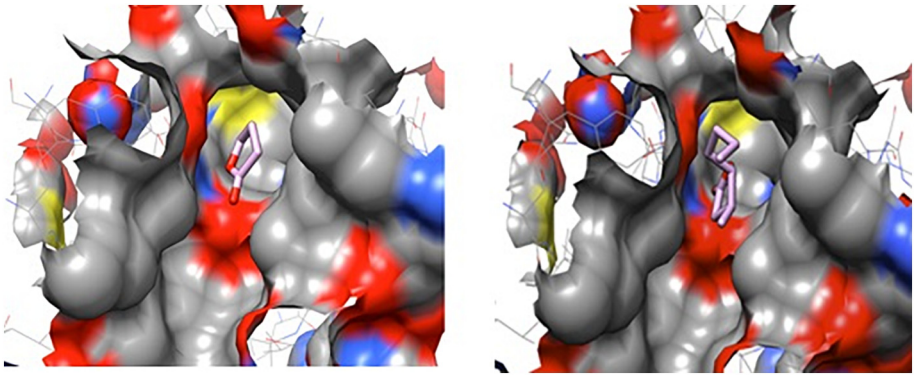


Fig. 8. Docking of substrates into the active site of the *C. hydrogenoformans* lactonase A) the GBL docked into the active site of the lactonase enzyme. B) a view of the (3*S*,4*S*)-WGL docked into the active site of the lactonase enzyme. The methyl group of the WGL prevents this substrate binding in the active conformation for catalysis and explains the experimental results showing no activity as on this substrate.

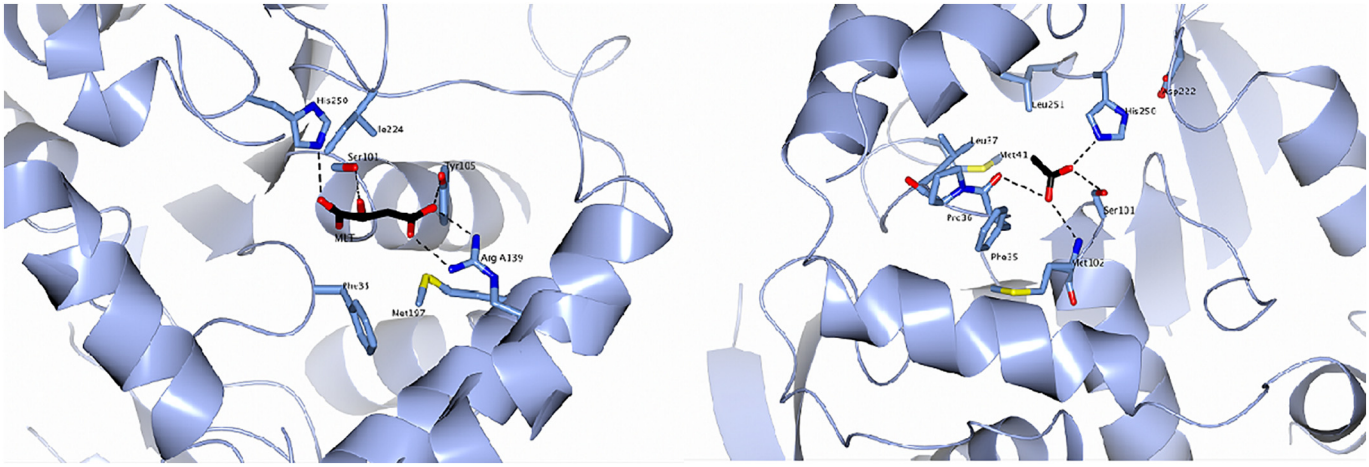


Fig. 9. Structure of carboxyl esterase TtEst (PDB: 4UHC) from *T. terrifontis*. A) The alcohol binding site of TtEst showing the protein backbone in cartoon mode. The important amino acid residues and the bound ligand are shown in stick mode. The dashed represent hydrogen bonds and ionic interactions. B) The carboxyl binding site which restricts the size of the substrate acyl group that can bind is shown. The bound acetate group and the important amino acid side chains are shown as stick models and the protein backbone in cartoon mode. The hydrogen bonds are shown as dashed lines.

the structural database shows its relation to 3-oxoadipate-enol lactonase from *Burkholderia xenovorans* Lb400 (PcaD; 29% sequence identity; PDB: 2XUA) (Bains et al., 2011), esterase from *Pseudomonas fluorescens* (PfEst; 30%; PDB: 3HEA) (Yin et al., 2010), (−) γ-lactamase from *Aureobacterium* sp. (Agl; 30%; PDB: 1HKH) (Line et al., 2004) and non-heme haloperoxidases from several bacteria including chloroperoxidase from *P. fluorescens* (PfCpo; 28%; PDB: 1A8S) (Hofmann et al., 1998).

The enzymes all show conservation of their positions of the active site catalytic triad and the the oxyanion hole. The type of catalytic reaction and the different substrate specificities are dictated by the residues of the small and large active site pockets. The TtEst binding pocket is larger than the equivalent PcaD and AgI pockets. The charged and much more polar nature allows D-malate and other organic acids to bind. The more hydrophobic PcaD and AgI binding pockets contain the bulky Trp135/Trp204 replacing Arg139 and the Ile129/ Leu125 in place of the Tyr105 in TtEst. The larger pocket is suitable for binding the cyclic lactones and γ-lactams in both of these enzymes. The hydrophobic side chains of the phenylalanines in the larger PfEst pocket able to bind a cyclic lactone would provide an explanation for its additional lactonase activity. The smaller TtEst pocket is designed for the small carboxyl ester groups to bind.

Another novel carboxyl esterase enzyme, TtEst2 from *T. terrifontis* was cloned, over-expressed, purified and crystallized with condition C6 of the Stura Footprint screen (Sayer et al., 2015b). These crystals belonged to space group $P2_12_12_1$ (cell dimensions a = 61.6, b = 71.0, c = 75.8 Å, α = β = γ = 90°) and diffracted to 1.6 Å. The TtEst2 structure showed the usual core domain of the α/β hydrolase fold but lacked most of the cap domain when compared to structurally related proteins. When compared with the esterase from *Alicyclobacillus acidocaldarius* (30% amino acid sequence identity) (De Simone et al., 2000) it was shown that this enzyme has an extra helix-loop-helix region at the N-terminus. The TtEst2 enzyme shows no contribution to the cap domain from the N-terminus. The *A. acidocaldarius* esterase has an extra loop-helix region in its cap domain that folds over the core domain which is not present in TtEst2 (shown in Fig. 10). This gives the TtEst2 a very open active site where the catalytic serine (Ser126) residue is positioned at the base of a groove that traverses the width of the enzyme (shown in Fig. 11).

5.4. An esterase enzyme (Est-2) from Archaeoglobus fulgidus

A putative esterase gene (Est-2) from the genome of *A. fulgidus* has been cloned, over-expressed and purified (Sayer et al., 2016). This enzyme Est-2 was characterized biochemically and structurally and showed esterase activity towards a range of *p*NP esters, with the butyrate and valerate esters being optimal substrates. Crystals of Est-2, obtained with the JCSG screen, belonged to spacegroup $P2_12_12_1$ (cell dimensions a = 55.6 Å, b = 67.6 Å, c = 139.5 Å, α = β = γ = 90°) and diffracted to 1.4 Å resolution. The structure was solved by molecular replacement using the *Mycobacterium tuberculosis* steroid-degrading hydrolase (PDB: 2VF2) as a model, with which it shares 27% sequence identity.

The Est-2 structure, showing the usual α/β hydrolase fold, consists of a core domain and a cap domain with Ser89, His228 and Asp200 making up the catalytic triad (Sayer et al., 2016). The carboxyl pocket of the Est-2 enzyme is defined by the Ser32, His88, Lys154 and Met231-side chains. These residues allow the *p*NP substrates of carbon chain lengths C1–5 to bind with optimal activity towards *p*NP-valerate, however there is not room for larger substrates to bind. The Est-2 structure (PDB: 5FRDFig. 12) is novel since it showed Coenzyme A (CoA) to be bound close to the enzyme's active site. The enzyme was demonstrated to have no CoA thioesterase activity, which has been described for a related human carboxyl esterase 1 (Bencharit et al., 2006). The structure of Est-2 shows that the alcohol pocket of the active site is partially obstructed by the CoA pantetheine group, which suggests that it may have a role in regulating its enzyme activity. Comparing the way how CoA binds to the human enzyme and Est-2 shows that it is binding in a different way, since CoA approaches the active site of these two enzymes from opposite directions.

The Est-2 has been shown to be active towards the industrially relevant substrate methyl *p*-toluate (Fig. 13). Initial rates of activity were measured by following the pH change upon *p*-toluic acid production using the indicator dye phenol red at 540 nm. The enzyme favours a neutral pH with an optimum of pH 9.0. Stability experiments showed Est-2 was able to withstand extreme conditions such pH 3.0 and 12.0 for 1 h, retaining over 88% relative activity. The Est-2 was very thermostable retaining its full activity after being incubated for 30 min at 75 °C (Sayer et al., 2016).

5.5. Other thermostable esterases

Metagenomic mining has also uncovered a new family of esterolytic enzymes of bacterial origin. A new esterase, termed EstDZ2, was discovered from a metagenome originating from a thermal spring in the Kamchatka Peninsula in the Far Eastern region of Russia (Menzel, et al.,

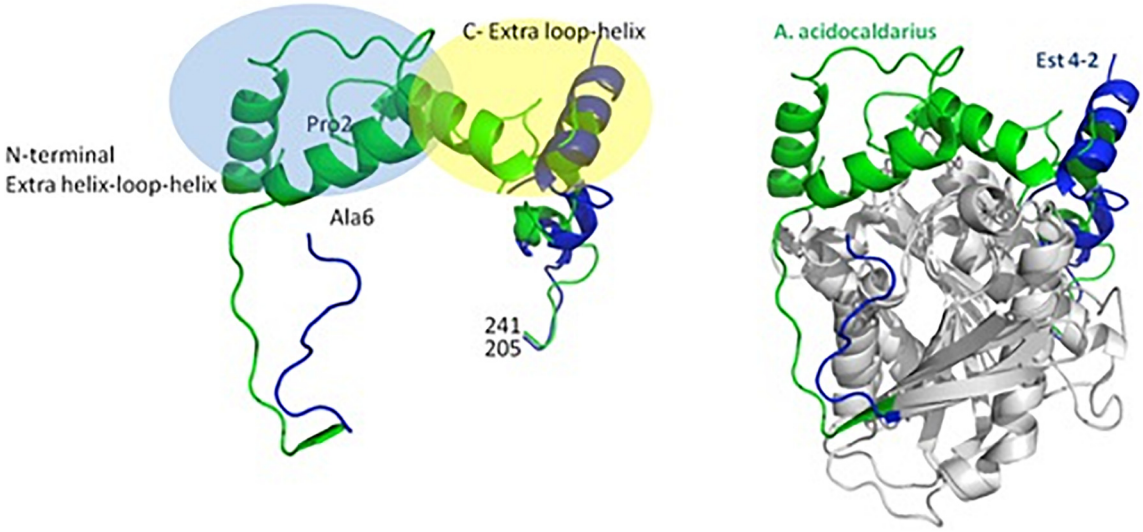


Fig. 10. A comparison of the carboxyl esterase TtEst2 (PDB: 5AO9) of *T. terrifontis* (blue) with the *A. acidocaldarius* esterase (PDB: 1EVQ) (green) highlighting the differences in the cap domain at the N-terminus and towards the C-terminal end of the enzyme. The lack of a cap domain in the TtEst2 enzyme means that this esterase has a very exposed active site compared to related esterases.The conserved core domain is shown in grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

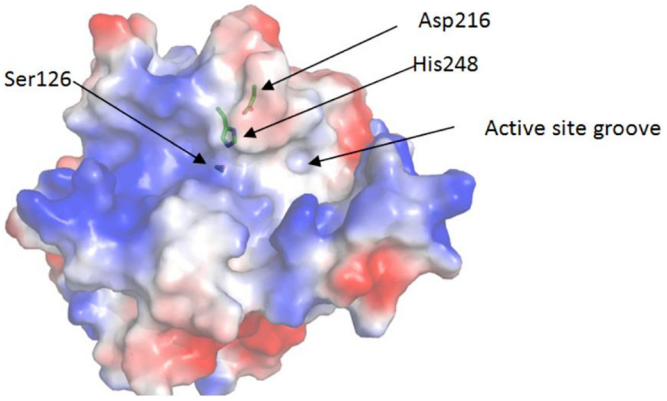


Fig. 11. An electrostatic surface representation of the carboxyl esterase TtEst2 of *T. terrifontis* with the esterase catalytic triad highlighted as stick models. This highlights the exposed active site groove running along the surface of the enzyme.

2015). Unbiased phylogenetic analysis demonstrated that the EstDZ2 amino acid sequence forms a deep branch that includes only functionally uncharacterized bacterial sequences, along with a few hydrolytic enzymes of animal and plant origin. This unique phylogenetic positioning suggested that EstDZ2 defines a novel family of esterolytic enzymes originating from bacteria (Zarafeta et al., 2016a), which has been named Family XV. EstDZ2 and other Family XV enzymes were found to be most closely related to Family IV and to a smaller extent to Family VII esterases. However, Family XV enzymes contain a highly conserved GHSAG catalytic motif, which is rare in other families of esterolytic enzymes. The nearest EstDZ2 homologue was found to be an uncharacterized lipase from "*Candidatus Acetothermus autotrophicum*" (Takami et al., 2012), a non-culturable representative of *Acetothermia* (OP1) candidate division (Hugenholtz et al., 1998), which is considered one of the deepest bacterial lineages (Jumas-Bilak et al., 2009). OP1 was one of the putative new bacteria phyla identified a few years ago following 16S rRNA sequencing of genetic material recovered from the Obsidian pool (OP), a thermal spring located at the Yellowstone National Park, USA (Rohini Kumar and Saravanan, 2010). It remains uncultured until today and poorly characterized, hence, a candidate division (phylum). Interestingly, all five open reading frames contained in the *estDZ2* contig encode for proteins whose nearest homologues are found in "*Candidatus Acetothermus autotrophicum*". This indicates that the entire contig originates from OP1. Hence, it appears that *estDZ2* has not been horizontally transferred recently, but has been evolutionary stable for a while, a feature supporting a functional role for the EstDZ2 esterase in this microorganism. In terms of catalytic properties, EstDZ2 efficiently catalyzed the hydrolytic cleavage of medium chain fatty acid esters at 25–60 °C and pH 7–8. Furthermore, EstDZ2 exhibited moderate thermostability (half-life of > 6 h at 60 °C), but possessed high stability at elevated concentrations of organic solvents, such as methanol, ethanol, acetone, isopropanol, 1-butanol, acetonitrile, isooctane and n-hexane (Zarafeta et al., 2016a). Multiple sequence alignment, combined with computational modelling of its 3D structure, revealed that EstDZ2 lacks the largest part of the "cap" domain, a structural domain that is encountered in all but a small number of esterolytic enzymes (Wei et al., 1998; Benavente et al., 2013; Sayer et al., 2015) and whose expanded structure characterizes the closely related Family IV esterases structurally (Wei et al., 1999; De Simone et al., 2000; Mandrich et al., 2005). These results suggest that EstDZ2 differs significantly from previously characterized esterolytic enzymes both in sequence and structure.

Other new thermostable esterase enzymes have been identified by bioinformatics analysis and these have provided knowledge of the increased diversity within this enzyme class. A very thermostable esterase has been identified in a microbial community isolated from a Chinese hot spring (Zarafeta et al., 2016b). Genomic screening identified it to originate from a *Dictyoglomus* bacterium. Biochemical characterization of this esterase revealed its preference for slightly basic conditions and fatty acid ester substrates with short to medium carbon chain length. The enzyme is remarkably thermostable and can maintain activity after exposure to 95 °C for several hours. Furthermore, it is very stable when exposed to high concentrations of organics. The so-called EstDZ3 is a member of the family 5 α/β hydrolases. Molecular modelling of the enzyme structure has revealed a subdomain insertion, which is also encountered in its closest homologue, the cinnamoyl esterase Lj0536 originating from *Lactobacillus johnsonii* (29% sequence identity). This subdomain insertion in the Lj0536 enzyme is proposed to be important for its substrate specificity (Lai et al., 2011).

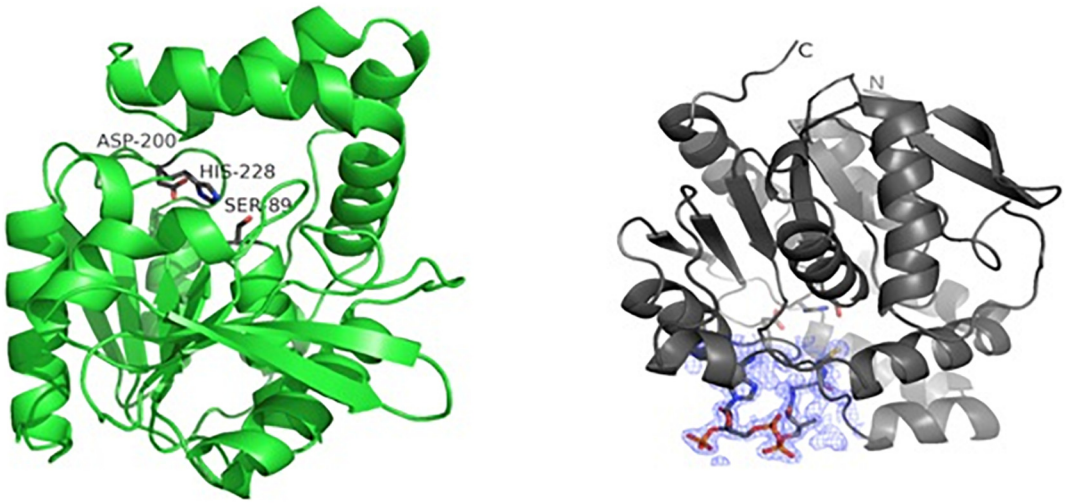


Fig. 12. The α/β hydrolase fold of the *A. fulgidus* Est-2 with the esterase catalytic triad shown as stick models (PDB code 5FRD). This enzyme has a coenzyme A molecule tightly bound at a unique position which is different from that seen for the thioesterases. The $2F_o$-F_c electron density map contoured at 1.2σ shows the presence of coenzyme A clearly in the protein structure bound close to the active site. The CoA molecule and active site residues are shown as stick models.

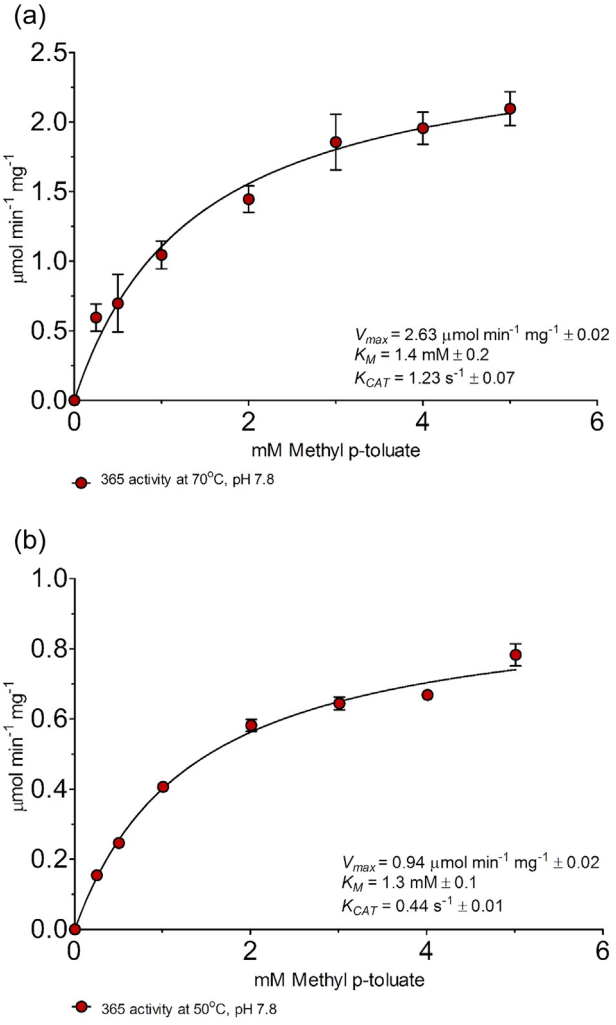


Fig. 13. A figure showing the kinetics of the reaction of the *A. fulgidus* Est-2 at 50 and 70 °C with the industrially interesting substrate, methyl *p*-toluate.

5.6. Novel limonene-1,2-epoxide hydrolases

A bioinformatics search of the HotZyme metagenomic assemblies available from the Anastasia platform was carried out to identify novel sequences showing similarity to the limonene-1,2-epoxide hydrolases (LEHs) from *Rhodococcus erythropolis* and *Mycobacterium tuberculosis*. Cloning and heterologous production of two newly identified LEHs (Tomsk-LEH and CH55-LEH from metagenomes isolated in Russia and China, respectively) was carried out together with that for the well studied *R. erythropolis* LEH (*Re*-LEH) to act as a control (Ferrandi et al., 2015b).

The LEH variants were obtained in good yields and purity by purification on a Ni-NTA resin and their performances under different reaction conditions were evaluated using the model substrate cyclohexene oxide and monitoring the biotransformations by chiral GC analysis. The influence of pH on the activity and stability of the novel EHs was evaluated within the pH range between pH 5.0 and 9.5 because most of the epoxide substrates are not stable at either very high or very low pHs. The best performances were observed at pH 8.0 for both enzymes. In agreement with their environmental origin, both enzymes showed a thermophilic character, with optimal activity at 40 °C and 60 °C for Tomsk-LEH and CH55-LEH, respectively. The observations recorded were consistent with the high melting temperatures of the two novel LEHs estimated by thermal shift analysis (74.5 °C and 79.7 °C for Tomsk-LEH and CH55-LEH, respectively).

Crystallization studies performed on the purified enzymes provided high quality crystals suitable for structure determination. The 3D structure of Tomsk-LEH was solved at 1.3 Å resolution (PBD: 5AIF) by molecular replacement using *Re*-LEH as a model. Subsequently, the CH55-LEH structure (PDB: 5AIH) was obtained starting from the corresponding crystals, which diffracted to 1.4 Å resolution, and solved using Tomsk LEH (48% sequence identity).

For both enzymes, the structural characterization confirmed the occurrence of a dimeric quaternary structure which is in agreement with the behaviour shown by the proteins during gel filtration chromatography under native conditions. Moreover, they showed to be very similar, from a structural point of view, to each other as well as to *Re*-LEH (Fig. 14A). In fact, the tertiary structure of each monomer consists of a six-stranded β-sheet, which is curved and mixed and where three α-helices are packed onto its concave side to form a deep active site cavity. More detailed analysis of the active site pocket showed the presence of a cluster of polar/charged residues (Asn34/34, Arg78/80, Asp111/112, and Asp80/82 in Tomsk-LEH and CH55-LEH,

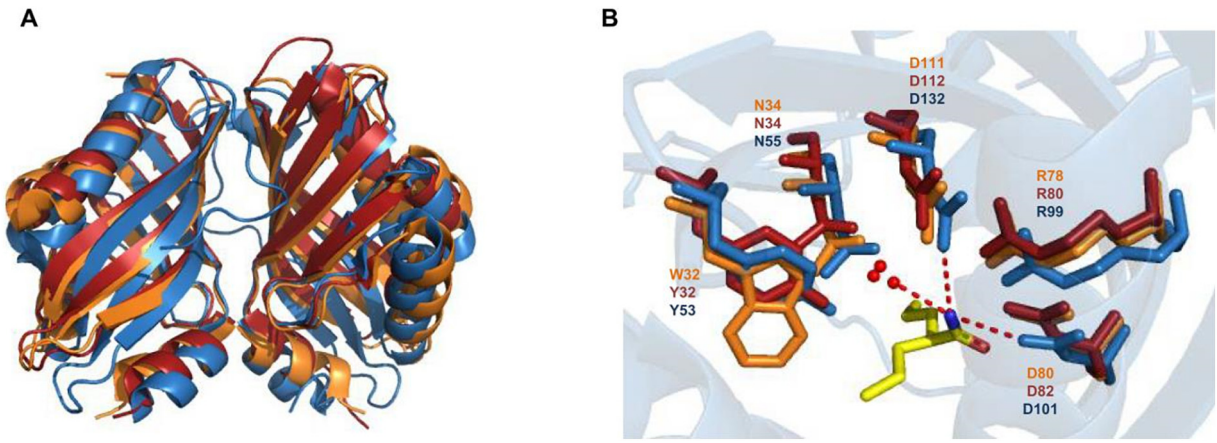


Fig. 14. Structure comparison of the thermostable limonene-1,2-epoxide hydrolases with the LEH of *R. erythropolis*, A) Overall structure alignment of Tomsk-LEH (PDB: 5AIF, *orange*), CH55-LEH (PDB: 5AIH, *red*), and *R. erythropolis* LEH (*Re*-LEH, PDB: 1NU3, *blue*). B) Key residue side chains in the active pocket (Tomsk-LEH, *orange*; CH55-LEH, *red*, and *Re*-LEH, *blue*). The inhibitor valpromide co-crystallized in the *Re*-LEH structure is coloured yellow (colour by element) which locates the potential position of substrate binding, while the catalytic water molecules are shown as red spheres. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

respectively), in a predominantly hydrophobic environment (Fig. 14B). The conservation of these residues in *Re*-LEH, where their essential role in the catalytic mechanism has been previously demonstrated, supports the fact that the discovered LEHs act in oxirane ring activation and hydrolysis by a similar way. Moreover, a co-crystallized complex of Tomsk-LEH crystals with the epoxide hydrolases inhibitor valpromide (PDB: 5AIG) showed hydrogen-bonding interactions between the carboxamide group of valpromide, Asp80 and a catalytic water molecule. However, the observed different positioning of α3 helix suggests the existence a slightly less accessible and more hydrophobic active site in the novel LEH enzymes than in *Re*-LEH. As far as the superior thermostability of the metagenome-derived LEHs, a comparative computational analysis has recently allowed the identification of key determinants ("hotspots") of LEHs structural stabilization (Rinaldi et al., 2017). Interestingly, while the mesophilic *Re*-LEH and the novel thermostable LEHs share a very similar flexibility profile with a main folding core in each protein monomer, both Tomsk-LEH and CH55-LEH showed an extended pattern of intra- and inter-monomer interactions that can have a relevant role in their stability at high temperatures. The stabilizing contribution of specific amino acidic residues was probed by structural characterization of variants obtained by site-directed mutagenesis.

The effects of the structural variations on the specific activity and selectivity of the LEH variants in the hydrolysis of different epoxide substrates, e.g., limonene-1,2-epoxides, cyclic *meso*-epoxides, aromatic and aliphatic terminal epoxides, were evaluated by estimating the performances of the enzymes in kinetic resolution and desymmetrization reactions (Ferrandi et al., 2015b).

Among the tested substrates, a higher activity of the novel LEHs was shown towards the *meso*-epoxide cyclohexene oxide, while lower conversion rates were observed with the C_5– or C_7-ring derivatives. The novel LEHs were generally significantly more stereoselective than *Re*-LEH. Moreover, the (1*S*,2*S*)-diol was always the preferred product of the thermostable variants, while *Re*-LEH showed a slight preference for the formation of the (1*R*,2*R*)-product in the case of the cyclopentene and cyclohexene epoxides. When tested in the hydrolysis of racemic mixtures of different terminal epoxides, CH55-LEH showed a high activity in the presence of the aromatic derivative styrene oxide, although with quite low enantioselectivity.

The most interesting results in terms of stereospecificity were obtained when comparing the activity of the thermostable LEHs with that of *Re*-LEH towards different limonene-1,2-epoxide isomers. In fact, differently from *Re*-LEH, both Tomsk-LEH and CH55-LEH preferred the *trans* isomer of (+)-limonene oxide and the *cis* isomer of (−)-limonene oxide (Fig. 15).

A preliminary structural analysis indicated that the minor accessibility of the active site in Tomsk-LEH and CH55-LEH when compared to *Re*-LEH may be related to the observed opposite stereopreference. In fact, in the absence of particular steric constraints, i.e., in the case of *Re*-LEH, the hydrolytic opening of the *cis* isomer of (+)-limonene oxide, as well as of the *trans* isomer of (−)-limonene oxide, is energetically favourable due to the presence and conformation of the isopropenyl substituent.

A more detailed structure-based rationalization of the observed opposite stereoselectivity was obtained by modelling the four limonene oxides isomers into the LEHs active sites. This study has allowed the identification of different steric factors which could limit the access of the limonene oxide isomers in Tomsk-LEH and CH55-LEH, thus governing the stereoselectivity of epoxide opening when using these enzymes. In particular, the presence of a bulky tryptophan residue (Trp 62/60) in both Tomsk-LEH and CH55-LEH, instead of a valine (Val83) as in *Re*-LEH, may hinder the efficient binding of all the isomers, as well as their interaction with catalytically relevant active site residues. This hypothesis was confirmed by studying the Trp62/60/Val variants of the Tomsk-LEH and CH55-LEH enzymes. In fact, in both cases, a significant change in the stereospecificity in limonene oxides opening was shown. The stereopreference of Tomsk-LEH W62 V for trans-(+)-limonene oxide and cis-(−)-limonene oxide decreased remarkably, while in the case of CH55-LEH W60 V, a stereospecificity switch towards the *cis* form of (+)-limonene oxide and a lack of preference for the hydrolysis of any of the two possible isomers of (−)-limonene oxide were observed.

Four complete sequences showing an α/β hydrolase fold with good homology with known or predicted EHs were identified by searching into the HotZyme assemblies. PCR amplification of the genes coding for the EH homologues from metagenomes was successfully performed and has enabled two new EHs which have been characterized biochemically and structurally (Ferrandi et al., 2018).

5.7. A novel cellulase

A novel cellulase was identified in the A5-7T enrichment sample through bioinformatics screening (Zarafeta et al., 2016c). The gene consists of 1155 bps and has a 59% sequence homology to a characterized endoglucanase from *Bacillus akibai*. Also, the sequence was 95% homologous to a putative glycoside hydrolase family protein from

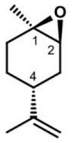
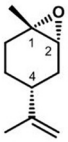
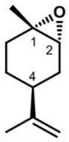
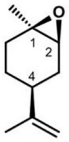
Enzyme	Specific activity (U/mg)			
	(+)-Limonene-1,2-oxide		(-)-Limonene-1,2-oxide	
	 (1*R*,2*S*,4*R*) (+)-*cis*	 (1*S*,2*R*,4*R*) (+)-*trans*	 (1*S*,2*R*,4*S*) (-)-*cis*	 (1*R*,2*S*,4*S*) (-)-*trans*
Re-LEH	30575	2400	1500	6475
Tomsk-LEH	22	200	150	4
CH55-LEH	293	440	365	71

Fig. 15. Specific activity of the thermostable limonene-1,2-epoxide hydrolases (LEHs) and *R. erythropolis* LEH (*Re*-LEH) towards different limonene oxides Isomers.

Thermoanaero-bacterium saccharolyticum. The protein was predicted to have a putative transmembrane helix at the N terminus, so the first 27 amino acids were replaced by a hexahistidine tag prior to recombinant production and purification. The truncated protein, designated as CelDZ1, showed improved expression levels and resistance to degradation compared to the full length protein. CelDZ1 was over-expressed from the pET-28a(+) expression vector in *E. coli* BL21 (DE3) and purified by immobilized metal ion chromatography. For biochemical characterization, pure enzyme was used in reactions containing CMC as the enzymatic substrate.

To obtain the pH profile of CelDZ1, the enzyme was assayed within the pH range of 4–10 at 40 °C. The assay revealed that pH 5 is the optimum pH for CelDZ1. At pH 6, the activity of the enzyme is 72% and at pH 4.9 and 10 the enzyme is inactive. CelDZ1 has a broad temperature range of action with activity of over 40% between 45 and 75 °C. Its optimum activity was at 70 °C and at 80 °C the activity drops to 21%, and is inactivated over 85 °C. The CelDZ1 remained stable when incubated at 65 and 70 °C for 24 h. When the enzyme was incubated at 75 °C, its activity dropped to 20% within the first hour of incubation. Importantly, CelDZ1 has been found to be highly halotolerant, with its activity remaining unchanged for up to 4 days in the presence of 5 M NaCl and 4 M KCl. This property of CelDZ1 is important for industrial processing of high salinity biomass, and unusual for an enzyme that was not isolated from a high-salt native environment.

The activity of CelDZ1 was also measured in the presence of a range of metal ions, reducing agents, detergents and organic solvents. $LiCl_2$, $CaCl_2$, $CuCl_2$ and $ZnCl_2$, when added at the concentration of 1 mM, had no effect on the enzyme, whereas $MnCl_2$ stimulated the enzyme's activity. Also non-ionic surfactants such as Triton X-100, Tween 20 and Tween 40 stimulated the enzyme when added at a concentration of 1% *v*/v. The addition of 1% of the organic solvents methanol and ethanol improved the activity of CelDZ1, 5% of the solvents caused a drop of the enzymatic activity down to 68% and 57% and at 10% of organic solvents concentration the activity of CelDZ1 reduced to 44% and 14%, respectively.

The substrate specificity of CelDZ1 was studied by assaying the enzyme against a range of polysaccharides. The substrates used in this assay included CMC, xylan, β-D-glucan from barley, cellobiose, avicel, galactomannan, laminarin and methyl cellulose. The CelDZ1 exhibits maximum activity against barley glucan followed by CMC. The rest of the substrates did not display measurable activity. Thus, it was concluded that CelDZ1 is a novel *endo*-glucanase for soluble cellulose. The Michaelis-Menten kinetic constants K_m and V_{max} were determined using the initial reaction rates for the different concentrations of CMC. The K_m was calculated at 22.73 g/l ± 3.24 and V_{max} at 0.41 ± 0.04 g/l/min.

CelDZ1 was crystallized in condition G6 of the Stura Footprint screen (Molecular Dimensions). Crystals diffracted to 1.9 Å in the space group C2 with unit cell parameters a = 107.3, b = 137.5, c = 121.3 Å, α = 90° β = 114° γ = 90°. The structure was solved using molecular replacement using the *Bacillus* alkaline cellulase K (PDB 1G01) (Zarafeta et al., 2016c). The structure of CelDZ1 is a conserved α/β barrel similar to other known GH5 structures. The catalytic residues were identified based on the sequence similarity and structural superimposition to the *Bacillus* enzyme. This identified Glu192 which functions as an acid/base and Glu294 which functions as a nucleophile in the catalytic mechanism. There is a clear groove running along one face of the enzyme and superimposition of the *Bacillus* enzyme shows the binding site of the cellobiose. The majority of the active site residues are conserved between the two enzymes with the only difference being Leu155 in the Cel5DZ1 structure compared to His333 in the *Bacillus* enzyme structure, which is in close proximity to the CH_2OH group of one of the glucose units. X-ray crystallography of CelDZ1 has revealed structural features which provide potential explanations about the biochemical characteristics of this new enzyme. CelDZ1 represents the first reported cellulase structure that lacks the defined sugar-binding 2 subsite which is present in other cellulases. While most of residues forming sugar binding subsites −3 to 1 are conserved between CelDZ1 and Cel5A (PDB 1H5V) the residues His206 and Gln 180 which form subsite 2 in Cel5A are replaced by Ala265 and Thr239 respectively in CelDZ1 and are unable to bind a sugar unit at this position (Fig. 16A and B). The highly negative surface of the enzyme especially in the active site channel is shown by the red colour (Fig. 17).

5.8. An unusual multi-domain glycosidase from Thermococcus sp.

For the identification of novel thermostable hydrolases, we applied an in-situ enrichment strategy (Kublanov et al., 2009, Gavrilov et al., 2016). 18 ml Hungate tubes were filled with the hot vent water and incubated in the respective hot vent, located in the tidal zone of Kunashir Island (Kuril archipelago). Xylan was added before incubation as a carbon and energy source. This enrichment strategy led to the identification of a novel hyperthermophilic member of the *Thermococcales* (Euryarchaeota), named *Thermococcus* sp. strain 2319 × 1. The organism showed optimal growth at 85 °C and neutral pH (pH 7.0) using a variety of polysaccharides such as amorphous cellulose (AMC), carboxymethyl cellulose (CMC), chitin, xyloglucan, and xylan.

The genome enabled the reconstruction of the different polysaccharide degradation pathways, identification of transport systems and central carbohydrate metabolism. Whereas the modified

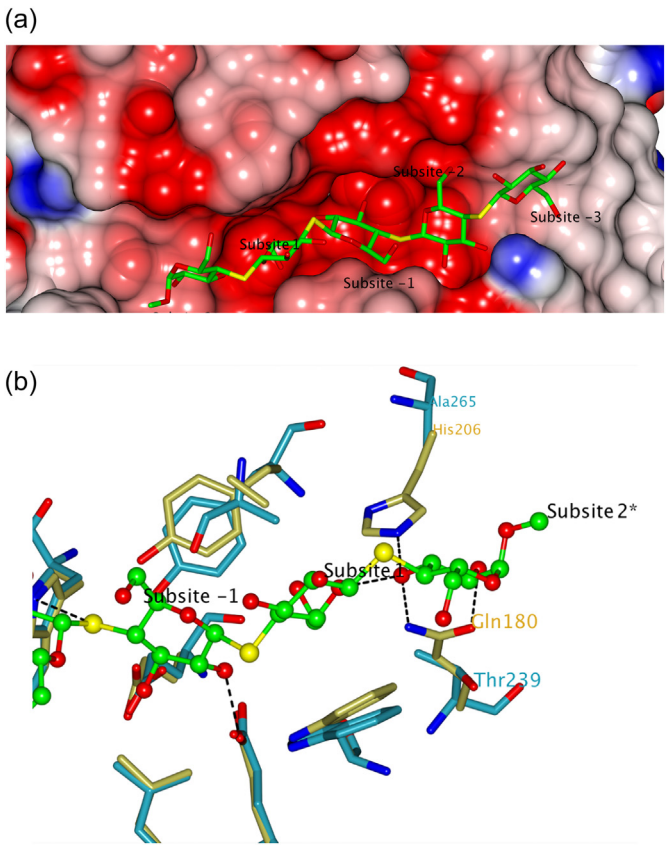


Fig. 16. A representation of the substrate binding in CelDZ1. A) The electrostatic potential surface of CelDZ1. The known structure of a related cellulase Cel5A (PDB 1H5V) with a bound substrate analogue, thiocellopentoside, was superimposed on CelDZ1. The bound ligand is shown as a stick model, the sugar binding subsites numbered. B) most of the residues forming sugar binding subsites −3 to 1 (ball and stick model) are conserved between CelDZ1 (cyan carbons) and Cel5A (yellow carbons) the residues His206 and Gln180 which form subsite 2 (marked by an asterisk) in Cel5A are replaced by Ala265 and Thr239 respectively in CelDZ1 and are unable to bind a sugar unit at this position. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

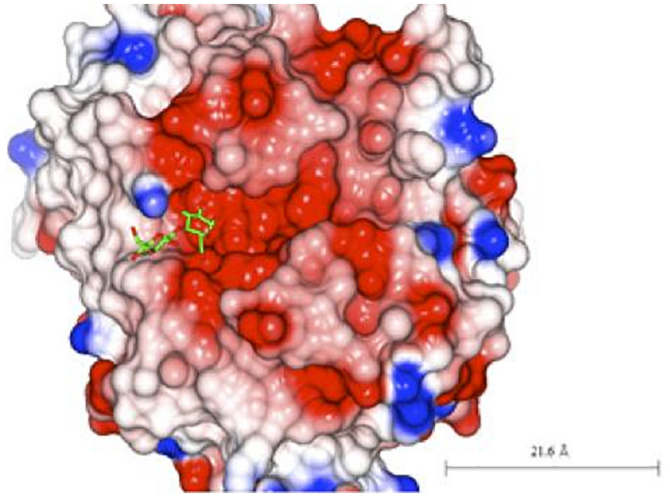


Fig. 17. A surface representation of the cellulase Cel5DZ1 enzyme. A superimposition with the *Bacillus* enzyme (PDB: 1G01) provides the location of the cellobiose binding site shown in green stick mode. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Thermococcales-like Emden-Meyerhof-Parnas pathway was identified for D-glucose degradation, the respective candidates involved in D-xylose metabolism could not be assigned pointing towards a novel metabolic route.

In addition, the available genome information allowed the identification of a novel multidomain glycosidase (MDG) with a new domain structure (Gavrilov et al., 2016). The enzyme comprises three glycoside hydrolase (GH) domains (one GH family 5 and two GH family 12 domains), and two carbohydrate-binding modules (family 2, CBM2). The respective domain order from the N- to C– terminus (GH5–12-12-CBM2–2) is shown in Fig. 18A. In order to retrieve functional information for the whole protein and different domains the full length MDG and different truncated versions with variable domain composition were heterologously expressed in *E. coli* and their enzymatic activity was analysed.

Among others the hydrolytic activity was analysed on carboxymethyl cellulose (CMC) agar plates (Fig. 18B). The MDG full length (~143 kDa) and truncated versions (i.e. GH5–12-12 (~125 kDa), GH5–12 (~88 kDa) and GH5 (~46 kDa)) were enriched by heat precipitation (60 °C) and 5–10 µg protein were spotted on CMC agar. As a negative control the crude extract of the expression strain (*E. coli* BL21 (DE3) pRIL) with empty vector (pET24a) was applied. Hydrolytic CMCase activity was confirmed for the MDG as well as for all truncated proteins on CMC agar plates (Fig. 18B).

In addition, the substrate specificity of the different recombinant proteins and their specific activity with different substrates was analysed. The full length MDG was found to hydrolyze various polysaccharides and displayed the highest activity for barley β-glucan (β-1,3/1,4-glucoside). Reduced activity was observed for carboxymethyl cellulose (β-1,4-glucoside), cello-oligosaccharides and galactomannan.

This approach embracing in situ enrichment, (comparative) genomics, cloning and classical biochemistry demonstrates the need and the potential of combined methods for the screening and characterization of novel enzymes with high biotechnological potential.

6. Industrial applications

Enzymes in their endogenous form have a deep association with mankind, with enzymes such as malting amylases from malted barley and rennin protease, isolated from stomach tissue for cheese production, to name two common examples. The modern era of enzymes and their industrial use however really only started during the 1960's when microorganisms, typically grown in large fermenters, became more widely used. The advantages of microbial production of enzymes are many and include: cheap and plentiful nutrition sources, rapid growth, high enzyme production, low biomass to enzyme output ratio and scalability. These advantages allow enzymes to be produced in quantities that enable them to be used in modern industrial processes (Aehle, 2007; Franssen et al., 2010; Buchholz et al., 2012; Grunwald, 2015; Liese et al. 2016; Faber, 2018) and to compete with traditional, often chemical based solutions, if indeed such solutions exist. The industrial enzyme market is estimated to be approximately 14.5 billion USD per year and growth is expected to continue.

Hydrolases have become workhorse catalysts for many different application areas like nutrition, bioenergy, chemical, agro, flavours and fragrances, cosmetic, biomedical and pharmaceutical industries (Meyer et al., 2013). The acceptance and inclusion of hydrolase-catalyzed reactions steps in industrial large-scale manufacturing processes has to do with the straightforward and well-established use of this robust and scalable enzyme technology requiring only standard equipment and no cofactors. Prerequisites for designing efficient and viable hydrolase-catalyzed industrial processes are clear targets for reaction performance, fast and meaningful analytical methods for measuring product formation and the availability of substrates and hydrolases with the desired properties. What makes hydrolases attractive for numerous applications in many different industrial areas is their diversity and

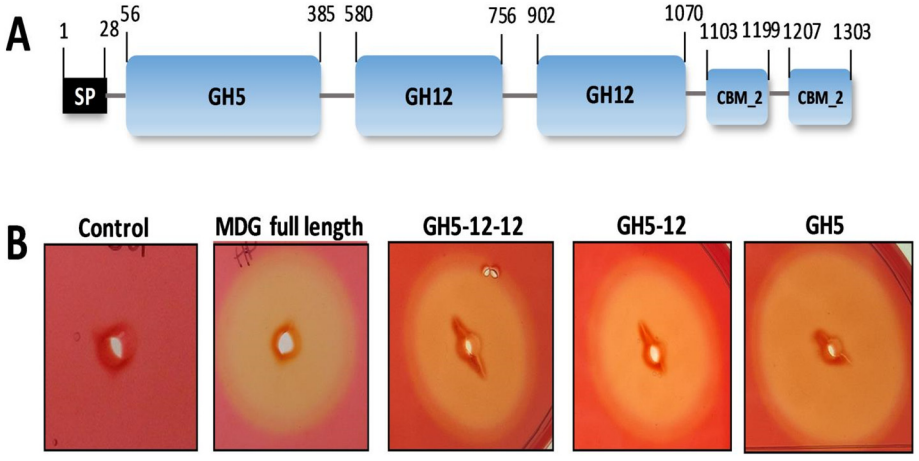


Fig. 18. Overview of the modular domain structure and cellulolytic activities of recombinant MDG and its truncated versions (modified from Gavrilov et al., 2016). Multidomain structure of the full length MDG (A), consisting of three glycoside hydrolase (GH) family domains and two family two carbohydrate binding modules (CBM2s). Domain order, N- to C-terminal direction: GH5-12-12-CBM2-2. (B) Hydrolytic activities of the full length MDG and truncated versions were analysed on CMC screening plates. Recombinant proteins were cloned without the signal peptide (SP) and the expression strain with empty plasmid was used as control. Plates were stained using 0.2% (w/v) Congo red and destained with 1 M NaCl, three times, 15 min at room temperature. For detailed discussion see text. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

versatility in hydrolysing a large range of substrate classes, from small to large molecular weight compounds.

The vast majority of industrial enzymes are hydrolases. These act on a variety of bonds, catalyze their cleavage and degrade compounds with high molecular weights into smaller components. This grouping includes protein-degrading enzymes. Proteinases have historically been widely used in detergents but new applications, such as feed enzymes have recently been established. By total sales, detergent enzymes is one of the largest segments. Savinase, a serine protease of the category PEPS8, is found in nearly all laundry detergent formulations. Proteases are effective for removing protein rich stains such as grass and blood from clothing. Besides enzymes like proteases or lipases in which the main effect is stain removal, enzymes can also act on the fabric itself. Certain cellulases that are active on amorphous cellulose can remove cellulose fibers and pills from the surface of cotton fabric while leaving the crystalline cellulose fibers constituting the fabric itself intact.

Another significant industrial enzyme class are amylases, which are used in two main areas: removal (desizing) of newly woven fabric and in the production of biofuels. In the first case, starch is used as a thread lubricant in high speed looms used to make the fabric. After weaving, the starch sizing agent must be removed and this is where amylase is used to hydrolyze the starch and therefore assist in this process. Amylases are also used in the total hydrolysis of starch to yield glucose which can be used directly or fermented into ethanol, the so named first generation ethanol production.

As opposed to first generation ethanol production, where essentially only amylases are needed to produce fermentable glucose, production of glucose from cellulosic biomass requires a suite of enzymes. This is because lignocellulose, as an important building block of the cell walls in plants, is resistant to degradation. An assortment of endo and exo acting cellulases are required to liberate the glucose disaccharide cellobiose, which is further degraded to glucose by glucosidase enzymes. One key enzyme in lignocellulose degradation for bioethanol applications is the cellulose bond breakage through the action of LPMO (lytic polysaccharide monooxygenase). LPMO enzymes can attack highly crystalline cellulose by oxidation at the carbon in the 1- or 4-position in cellulose (Hemsworth et al. 2013).

Novel hydrolases have also shaped very much the synthesis of a large number of chiral and non-chiral small molecular weight compounds like amides, esters, lactones and epoxides. Hydrolases have also been shown to catalyze the reverse reaction of synthesis by utilizing various tools for avoiding the hydrolysis direction, such as the use of organic solvents, irreversible donors or specifically selected or engineered hydrolases. The ability to dissolve water-insoluble substrates/products, to change the thermodynamics of the reaction system or to increase selectivity can be an additional advantage in using non-conventional media like organic solvents. Even though hydrolase-catalyzed reactions are preferred reaction steps, their early integration into the overall synthetic route is essential, e.g. when resolutions, deracemizations, desymmetrizations or deprotections are considered. Hydrolytic resolutions can build on a large variety of easily accessible substrates, while desymmetrizations of prochiral substrates are attractive due to their potential of complete conversion into one enantiomerically pure product. The translation from laboratory to industrial large scale can build on a vast experience in hydrolase-catalyzed reactions (Blaser and Federsel, 2010), from engineering of batch or continuous reactions, use of soluble or immobilized enzymes, aqueous or non-aqueous media to downstream processing, product recovery and purification.

The synthetic application of epoxide hydrolases with extraordinarily high stereoselectivity in numerous biocatalytic epoxide ring-opening reactions aimed at the preparation of chiral epoxides and diols has been widely demonstrated in the last years (Kotik et al., 2012; Wohlgemuth, 2015). Recently, a set of different limonene epoxide hydrolases (LEHs) showing complementary stereopreference for the limonene oxide isomers has been successfully employed in the enzymatic preparation of all limonene oxide stereoisomers and limonene diols in enantiomerically pure form (Ferrandi et al., 2015a). Fine-tuning of the biotransformation processes allowed the simultaneous gram-scale preparation of different limonene oxides and diols in high enantiomeric purity by choosing the biocatalyst with the desired stereospecificity (Fig. 19). It is worth mentioning that all these epoxide resolutions could be achieved with good to excellent space-time yields and productivity under solvent-free conditions as all the LEHs were capable to act in a biphasic reaction system with neat substrates as a water-immiscible phase.

Beside the importance of individual structural and molecular characteristics of hydrolases and their catalytic mechanisms for the hydrolysis of large molecular weight substrates as well as small molecular weight compounds, there are also common aspects in industrial applications. The common features about the many different novel thermostable hydrolase activities and their specific industrial applications are listed in Table 3.

The integration of desired targets for products and hydrolase-catalyzed bioprocesses with the search and selection of the ideal hydrolases, the process design, reaction engineering and rapid prototyping is key for the many successful applications in numerous industries (Ghisalba et al. 2010; Meyer et al. 2013; Wohlgemuth, 2017; de Gonzalo and de Maria, 2018).

7. Perspectives

For the transition towards more sustainable processes, we need novel robust biocatalysts that can resist extremes of pH and temperature, solvents and salts. Extremophilic microbes are a natural reservoir of such robust enzymes. While the HotZyme project contributed to this

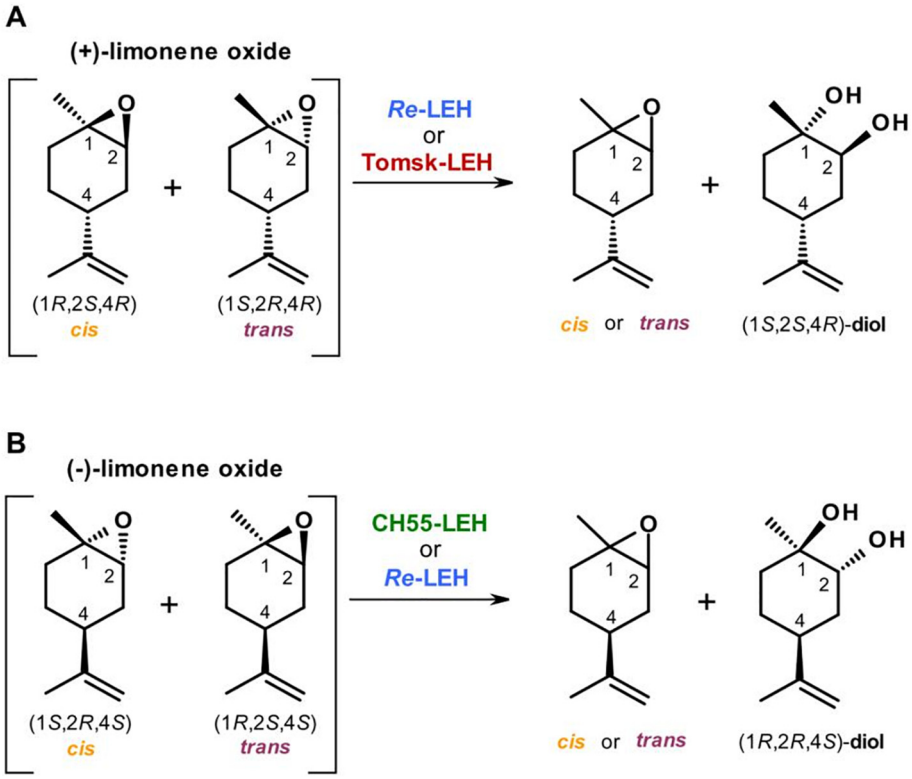


Fig. 19. Gram-scale kinetic resolutions of (+)- and (−)-limonene oxide (A and B, respectively), catalyzed by LEHs showing complementary stereo-preference.

goal by identifying novel thermostable hydrolases from high temperature microbiomes, many challenges remain in all aspects of such an undertaking.

The mining of high temperature environments remains a rich source of new thermostable enzymes and metabolites and their potential is promising and just beginning to be explored. Advances over the past decades of various 'omics' technologies - such as genomics, transcriptomics, proteomics, or metabolomics - have made their application to diverse high temperature environments feasible for life science laboratories. Generating enormous environmental datasets is not the limitation, however, mining these large and complex datasets for relevant targets is still challenging. The greatest challenge is gene annotation, especially of the large number of genes that have no obvious function. Due to errors in databases, also regarding inference of gene functions by sequence identity, annotation might not give the actual function. Sequence annotation is therefore only a first step towards functional characterization. As the HotZyme Project has demonstrated, it takes a combination of approaches to succeed. In the future, continued efforts at understanding the fundamental biochemistry of microbes from high temperature environments will undoubtedly advance our ability to mine these environments for useful products.

For functional screening, an enormous amount of different screening techniques are available, but unfortunately they are not yet applicable for every individual enzyme. The major challenges for each screening method are the heterologous expression of the enzyme, the possibility to screen in a high-throughput manner and the substrate availability for both the researcher and the enzyme. Expanding the host repertoire, using high-throughput bioreporter-based or microfluidics-based screening and developing new substrate synthesis methods could help tackle most of these challenges. Although in vivo screening always faces the challenge of getting the substrate and the enzyme together due to the presence of the cell membrane, there are ways around this, for example by modification of the cell membrane to incorporate additional compounds and by cell lysis and using an in vitro system Once target enzymes have been identified and characterized, a synthetic biology approach can be developed for in vivo production of important molecules by introduction of non-natural pathways in an appropriate host organism such as *E. coli*. New non-natural pathways can also be developed in vitro by performing enzymatic reactions which are compatible with each other in one pot or separating in space or time

Table 3
Common features and specific applications of the novel thermostable hydrolases.

Enzyme activities	Common features and specific applications
Lactonases	Synthesis of γ-hydroxycarboxylic acids and resolution of racemic lactones by selective hydrolysis of lactone rings
Esterases	Hydrolysis of small carboxylic esters and γ-lactams, resolution of racemates, high stability at elevated concentrations of organic solvents
Lipases	Hydrolysis of large carboxylic esters, resolution of racemates, stain removal by fat and lipid hydrolysis
Epoxide Hydrolases	Synthesis of epoxides and vicinal diols, resolution of mixtures of cis/trans-isomers by selective ring-opening of epoxides
Cellulases	Hydrolysis of cellulose in biomass conversion and textile treatment
Xylanases	Hydrolysis of xylan polysaccharides in pulp and paper applications and biomass conversion
Proteases	Stain removal by protein hydrolysis in detergent and laundry applications
Cutinases	Degradation of polyethylene terephthalate
Lignin degrading enzymes	Depolymerization of lignin in biomass conversion, pulp and paper applications
Glycosidase-Endoglucanase/ Mannosidase	Hydrolysis of β-glucan

enzymatic reactions which are incompatible (Wohlgemuth, 2018).

As it is also important to describe and store the wealth and growing understanding of novel hydrolase structure-function relationships in a way which allows fast retrieval, novel bioinformatics tools and the enzyme function database STRENDA (Gardossi et al., 2010; Swainston et al., 2018) will contribute to further expand applications of hydrolases which have been already discovered.

8. Conclusions

The search for thermostable hydrolases with novel functions and appropriate performance for industrial applications is of much interest, because thermostable enzymes provide additional benefits compared to their mesophilic counterparts. These benefits include higher stability and better productivity. However, discovering the desired hydrolases presents a lot of challenges ranging from microbial sampling to characterization. The Hotzyme project therefore aimed to discover novel thermostable hydrolases from hot environments by using a multidisciplinary approach that combined various mining and screening methods, bioinformatics, and detailed functional and structural characterization of the discovered enzymes. Targeted enzyme classes included lactonases, esterases, epoxide hydrolases, cellulases and xylanases. The fruitful outcome of this project comprised hundreds of environmental samples from hot springs worldwide, thousands of enrichment cultures growing on polymeric substrates of industrial interest, 15 hot spring metagenomes, several sequenced isolate genomes and transcriptomes, and 100 potentially new hydrolases of which more than a dozen were biochemically and structurally characterized. The characterized hydrolases include thermostable carboxylesterases, enol lactonases, quorum sensing lactonases, gluconolactonases, limonene epoxide hydrolases, αβ hydrolase fold epoxide hydrolases and cellulases. Apart from the novel thermostable hydrolases and the enormous potential to discover more of these enzymes from the obtained samples and data, several tools were developed to facilitate annotation and screening, for example the annotation platform Anastasia and a whole-cell bioreporter-based functional screening method. The success of the HotZyme project shows that there are still a lot of hydrolases to be discovered. It was especially the multidisciplinary approach, in which 13 groups each having a different expertise worked together in a consortium, that made this project so successful. The vast diversity of hydrolase enzymes from hot environments, the new approaches and tools developed provide an enormous treasure for expanding the already well established hydrolase applications and will be a very valuable asset for the further exploration of uncharted novel hydrolase territories.

Acknowledgements

We thank Celeste Annovazzi, Prof. Zohar Ben-Asher, Dr. Tom Bisschops, Prof. Dr. Elizaveta A. Bonch-Osmolovskaya, Dr. Alina Bunescu, Dr. Aristotelis Chatziioannou, Sjoerd C.A. Creutzburg, Dr. Erica E. Ferrandi, Dr. Sergey N. Gavrilov, Haimo Gilbert, Dr. Sóley Ruth Guðbergsdóttir, Soren Haslund, Dr. Colin J. Ingham, Dr. Michail N. Isupov, Dr. Kenneth Jensen, Dr. Verena Kallnik, Dr. Servé W.M. Kengen, Prof. Fragiskos N. Kolisis, Dr. Theresa Kouril, Dr. Anders Krogh, Dr. Efthymios Ladoukakis, Carlotta Marchesi, Saideh Mirlashari, Dr. Sarah Müller, Myhanh Nguyen, Prof. John van der Oost, Dr. Christopher Sayer, Dr. Vahid Saneii, Dr. Tatyana Sokolova, Christina Stracke, Aaron Westlake, Dr. Dimitra Zarafeta and all members of the HotZyme consortium for their great cooperation and their concerted efforts towards the discovery of novel hydrolases from hot environments.

Financial support from the European Union 7th Framework Programme FP7/2007-2013 Project HotZyme under grant agreement nr. 265933 is gratefully acknowledged. Research conducted in Yellowstone National Park USA was performed under Yellowstone Research Permit YELL-20XX-SCI-5090.

References

Aehle, W. (Ed.), 2007. Enzymes in Industry – Production and Applications, Third, completely revised edition. Wiley-VCH, Weinheim, Germany.
Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A., Goecks, J., 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 44, W3–W10.
Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
Bains, J., Kaufman, L., Farnell, B., Boulanger, M.J., 2011. A product analog bound form of 3-oxoadipate-enol-lactonase (PcaD) reveals a multifunctional role for the divergent cap domain. J. Mol. Biol. 406, 649–658.
Bankevich, A., Nurk, S., Antipov, D., Gurevich, A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477.
Baret, J.-C., Miller, O.J., Taly, V., Ryckelynck, M., El-Harrak, A., Frenz, L., Rick, C., Samuels, M.L., Hutchison, J.B., Agresti, J.J., Link, D.R., Weitz, D.A., Griffiths, A.D., 2009. Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity. Lab Chip 9, 1850.
Becker, F., Schnorr, K., Wilting, R., Tolstrup, N., Bendtsen, J.D., Olsen, P.B., 2004. Development of in vitro transposon assisted signal sequence trapping and its use in screening *Bacillus halodurans* C125 and *Sulfolobus solfataricus* P2 gene libraries. J. Microbiol. Methods 57, 123–133.
Benavente, R., Esteban-Torres, M., Acebrón, I., de Las Rivas, B., Muñoz, R., Alvarez, Y., Mancheño, J.M., 2013. Structure, biochemical characterization and analysis of the pleomorphism of carboxylesterase Cest-2923 from *Lactobacillus plantarum* WCFS1. FEBS J. 280, 6658–6671.
Bencharit, S., Edwards, C.C., Morton, C.L., Howard-Williams, E.L., Kuhn, P., Potter, P.M., Matthew, R., Redinbo, M.R., 2006. Multisite Promiscuity in the Processing of Endogenous Substrates by Human Carboxylesterase 1. J. Mol. Biol. 363, 201–214.
Blaser, H.-U., Federsel, H.-J. (Eds.), 2010. Asymmetric Catalysis on Industrial Scale, Second edition. Wiley-VCH, Weinheim, Germany.
Bolduc, B., Shaughnessy, D.P., Wolf, Y.I., Koonin, E.V., Roberto, F.F., Young, M., 2012. Identification of novel positive-strand RNA viruses by metagenomic analysis of Archaea-dominated Yellowstone hot springs. J. Virol. 86, 5562–5573.
Bolduc, B., Wirth, J.F., Mazurie, A., Young, M.J., 2015. Viral assemblage composition in Yellowstone acidic hot springs assessed by network analysis. ISME J. 9, 2162–2177.
Boock, J.T., Gupta, A., Prather, K.L.J., 2015. Screening and modular design for metabolic pathway optimization. Curr. Opin. Biotechnol. 36, 189–198.
Bornscheuer, U.T., Kazlauskas, R.J., 2006. Hydrolases in Organic Synthesis: Regio- and StereosSelective Biotransformations. John Wiley & Sons, New York.
Bornscheuer, U.T., Huisman, G.W., Kazlauskas, R.J., Lutz, S., Moore, J.C., Robins, K., 2012. Engineering the third wave of biocatalysis. Nature 485, 185–194.
Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A., Wilkins, M.J., Wrighton, K.C., Williams, K.H., Banfield, J.F., 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. Nature 523, 208–211.
Buchfink, B., Xie, C., Huson, D.H., 2014. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60.
Buchholz, K., Kasche, V., Bornscheuer, U.T., 2012. Biocatalysts and Enzyme Technology, Second, completely revised and enlarged edition. Wiley-VCH, Weinheim, Germany.
Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: Architecture and applications. BMC Bioinformatics 10, 1–9.
Catón, L., Yurkov, A., Giesbers, M., Dijksterhuis, J., Ingham, C.J., 2017. Physically triggered morphology changes in a novel *Acremonium* isolate cultivated in precisely engineered microfabricated environments. Front. Microbiol. 8, 1269.
Cedrone, F., Bhatnagar, T., Baratti, J.C., 2005. Colorimetric assays for quantitative analysis and screening of epoxide hydrolase activity. Biotechnol. Lett. 27, 1921–1927.
Choi, S.L., Rha, E., Lee, S.J., Kim, H., Kwon, K., Jeong, Y.S., Rhee, Y.H., Song, J.J., Kim, H.S., Lee, S.G., 2014. Toward a generalized and high-throughput enzyme screening system based on artificial genetic circuits. ACS Synth. Biol. 3, 163–171.
Cleary, B., Brito, I.L., Huang, K., Gevers, D., Shea, T., Young, S., Alm, E.J., 2015. Detection of low-abundance bacterial strains in metagenomic datasets by eigengenome partitioning. Nat. Biotechnol. 33, 1053–1060.
Colin, P.Y., Kintses, B., Gielen, F., Miton, C.M., Fischer, G., Mohamed, M.F., Hyvönen, M., Morgavi, D.P., Janssen, D.B., Hollfelder, F., 2015. Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. Nat. Commun. 6, 10008.
Colman, D.R., Feyhl-Buska, J., Fecteau, K.M., Xu, H., Shock, E.L., Boyd, E.S., 2016. Ecological differentiation in planktonic and sediment-associated chemotrophic microbial populations in Yellowstone hot springs. FEMS Microbiol. Ecol. 92, 1–13.
Daas, M.J.A., Nijsse, B., van de Weijer, A.H.P., Groenendaal, B.W.A.J., Janssen, F., van der Oost, J., van Kranenburg, R., 2018. Engineering *Geobacillus thermodenitrificans* to introduce cellulolytic activity: expression of native and heterologous cellulose genes. BMC Biotechnol. 18, 42.
De Gonzalo, G., de Maria, P.D., 2018. Biocatalysis : An Industrial Perspective. The Royal Society of Chemistry, London.
de los Santos, E.L.C., Meyerowitz, J.T., Mayo, S.L., Murray, R.M., 2016. Engineering transcriptional regulator effector specificity using computational design and in vitro rapid prototyping: developing a vanillin sensor. ACS Synth. Biol. 5, 287–295.
De Simone, G., Galdiero, S., Manco, G., Lang, D., Rossi, M., Pedone, C., 2000. A snapshot

of a transition state analogue of a novel thermophilic esterase belonging to the subfamily of mammalian hormone-sensitive lipase. J. Mol. Biol. 303, 761–771.
Eddy, S.R., 2011. Accelerated profile HMM searches. PLoS Comput. Biol. 7, e1002195.
Eggeling, L., Bott, M., Marienhagen, J., 2015. Novel screening methods - biosensors. Curr. Opin. Biotechnol. 35, 30–36.
Elcheninov, A.G., Menzel, P., Gudbergsdottir, S.R., Slesarev, A.I., Kadnikov, V.V., Krogh, A., Bonch-Osmolovskaya, E.A., Peng, X., Kublanov, I.V., 2017. Sugar metabolism of the first thermophilic planctomycete *Thermogutta terrifontis*: comparative genomic and transcriptomic approaches. Front. Microbiol. 8, 2140.
Elleuche, S., Schäfers, C., Blank, S., Schröder, C., Antranikian, G., 2015. Exploration of extremophiles for high temperature biotechnological processes. Curr. Opin. Microbiol. 25, 113–119.
Eloe-Fadrosh, E.A., Paez-Espino, D., Jarett, J., Dunfield, P.F., Hedlund, B.P., Dekas, A.E., Grasby, S.E., Brady, A.L., Dong, H., Briggs, B.R., Li, W.J., Goudeau, D., Malmstrom, R., Pati, A., Pett-Ridge, J., Rubin, E.M., Woyke, T., Kyrpides, N.C., Ivanova, N.N., 2016. Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. Nat. Commun. 7, 1–10.
Faber, K., 2018. Biotransformations in Organic Chemistry. Springer-Verlag, Berlin, Germany.
Ferrandi, E.E., Marchesi, C., Annovazzi, C., Riva, S., Monti, D., Wohlgemuth, R., 2015a. Efficient epoxide hydrolase catalyzed resolutions of (+)- and (−)- *cis/trans*-limonene oxides. ChemCatChem 7, 3171–3178.
Ferrandi, E.E., Sayer, C., Isupov, M.N., Annovazzi, C., Marchesi, C., Iacobone, G., Peng, X., Bonch-Osmolovskaya, E., Wohlgemuth, R., Littlechild, J.A., Monti, D., 2015b. Discovery and characterization of thermophilic limonene-1,2-epoxide hydrolases from hot spring metagenomic libraries. FEBS J. 282, 2879–2894.
Ferrandi, E.E., Sayer, C., Rose, S.A., Guazzelli, E., Marchesi, C., Saneei, V., Isupov, M.N., Littlechild, J.A., Monti, D., 2018. New Thermophilic α/β Class Epoxide Hydrolases Found in Metagenomes from Hot Environments. Front. Bioeng. Biotechnol. 6, 144.
Ferrer, M., Martinez-Martinez, M., Bargiela, R., Streit, W.R., Golyshina, O.V., Golyshin, P.N., 2016. Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. Microb. Biotechnol. 9, 22–34.
Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M., 2014. Pfam: the protein families database. Nucleic Acids Res. 42, 222–230.
Franssen, M.C., Kircher, M., Wohlgemuth, R., 2010. Industrial biotechnology in the chemical and pharmaceutical industries. In: Soetaert, W., Vandamme, E.J. (Eds.), Industrial Biotechnology, Sustainable Growth and Economic Success. Wiley-VCH, Weinheim, Germany.
Gardossi, L., Poulsen, P.B., Ballesteros, A., Hult, K., Švedas, V.K., Vasić-Rački, Đ., Carrea, G., Magnusson, A., Schmid, A., Wohlgemuth, R., Halling, P.J., 2010. Guidelines for reporting of biocatalytic reactions. Trends Biotechnol. 28, 171–180.
Gavrilov, S.N., Stracke, C., Jensen, K., Menzel, P., Kallnik, V., Slesarev, A., Sokolova, T., Zayulina, K., Bräsen, C., Bonch-Osmolovskaya, E.A., Peng, X., Kublanov, I.V., Siebers, B., 2016. Isolation and characterization of the first xylanolytic hyperthermophilic euryarchaeon *Thermococcus* sp. strain 2319x1 and its unusual multidomain glycosidase. Front. Microbiol. 7, 552.
Ghisalba, O., Meyer, H.P., Wohlgemuth, R., 2010. Industrial Biotransformation. In: Flickinger, M.C. (Ed.), Encyclopedia of Industrial Biotechnology. Wiley, Hoboken, NJ.
Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. Nat. Rev. Genet. 17, 333–351.
Grunwald, P. (Ed.), 2015. Industrial Biocatalysis. CRC Press Taylor & Francis Group, Boca Raton FL, USA.
Gudbergsdóttir, S.R., Menzel, P., Krogh, A., Young, M., Peng, X., 2016. Novel viral genomes identified from six metagenomes reveal wide distribution of archaeal viruses and high viral diversity in terrestrial hot springs. Environ. Microbiol. 18, 863–874.
Hauser, M., Steinegger, M., Söding, J., 2016. MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. Bioinformatics 32, 1323–1330.
Hemsworth, G.R., Davies, G.J., Walton, P.H., 2013. Recent insights into copper-containing lytic polysaccharide mono-oxygenases. Curr. Opin. Struct. Biol. 23, 660–668.
Hiblot, J., Gotthard, G., Chabriere, E., Elias, M., 2012. Structural and enzymatic characterization of the lactonase SisLac from *Sulfolobus islandicus*. PLoS One 7, e47028.
Hiblot, J., Bzdrenga, J., Champion, C., Chabriere, E., Elias, M., 2015. Crystal structure of VmoLac, a tentative quorum quenching lactonase from the extremophilic crenarchaeon *Vulcanisaeta moutnovskia*. Sci. Rep. 5, 8372.
Hoff, T., Schnorr, K.M., Meyer, C., Caboche, M., 1995. Isolation of two Arabidopsis cDNAs involved in early steps of molybdenum cofactor biosynthesis by functional complementation of *Escherichia coli* mutants. J. Biol. Chem. 270, 6100–6107.
Hofmann, B., Tölzer, S., Pelletier, I., Altenbuchner, J., van Pée, K.H., Hecht, H.J., 1998. Structural investigation of the cofactor-free chloroperoxidases. J. Mol. Biol. 279, 889–900.
Hugenholtz, P., Goebel, B.M., Pace, N.R., 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol. 180, 4765–4774.
Huson, D., Mitra, S., Ruscheweyh, H., 2011. Integrative analysis of environmental sequences using MEGAN4. Genome Res. 21, 1552–1560.
Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119.
Hynninen, A., Virta, M., Scheper, T., 2010. Whole-cell bioreporters for the detection of bioavailable metals. In: Belkin, S., Gu, M.B. (Eds.), Whole Cell Sensing System II. Springer Berlin Heidelberg, pp. 31–63.
Ingham, C.J., Sprenkels, A., Bomer, J., Molenaar, D., van den Berg, A., van Hylckama Vlieg, J.E.T., de Vos, W.M., 2007. The micro-Petri dish, a million-well growth chip for the culture and high-throughput screening of microorganisms. Proc. Natl. Acad. Sci. 104, 18217–18222.
Inskeep, W.P., Rusch, D.B., Jay, Z.J., Herrgard, M.J., Kozubal, M.A., Richardson, T.H., Macur, R.E., Hamamura, N., Jennings, R. de M., Fouke, B.W., Reysenbach, A.L., Roberto, F., Young, M., Schwartz, A., Boyd, E.S., Badger, J.H., Mathur, E.J., Ortmann, A.C., Bateson, M., Geesey, G., Frazier, M., 2010. Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. PLoS One 5, e9773.
Inskeep, W.P., Jay, Z.J., Tringe, S.G., Herrgård, M.J., Rusch, D.B., 2013. The YNP metagenome project: Environmental parameters responsible for microbial distribution in the yellowstone geothermal ecosystem. Front. Microbiol. 4, 1–15.
Jeong, Y.S., Choi, S.L., Kyeong, H.H., Kim, J.H., Kim, E.J., Pan, J.G., Rha, E., Song, J.J., Lee, S.G., Kim, H.S., 2012. High-throughput screening system based on phenolics-responsive transcription activator for directed evolution of organophosphate-degrading enzymes. Protein Eng. Des. Sel. 25, 725–731.
Jha, R.K., Kern, T.L., Kim, Y., Tesar, C., Jedrzejczak, R., Joachimiak, A., Charlie, E.M., Strauss, C.E.M., 2016. A microbial sensor for organophosphate hydrolysis exploiting an engineered specificity switch in a transcription factor. Nucleic Acids Res. 44, 8490–8500.
Jia, B., Raphenya, A.R., Alcock, B., Waglechner, N., Guo, P., Tsang, K.K., Lago, B.A., Dave, B.M., Pereira, S., Sharma, A.N., Doshi, S., Courtot, M., Lo, R., Williams, L.E., Frye, J.G., Elsayegh, T., Sardar, D., Westman, E.L., Pawlowski, A.C., Johnson, T.A., Brinkman, F.S.L., Wright, G.D., McArthur, A.G., 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res. 45, D566–D573.
Jumas-Bilak, E., Roudiere, L., Marchandin, H., 2009. Description of 'Synergistetes' phyl. Nov. and emended description of the phylum 'Deferribacteres' and of the family Syntrophomonadaceae, phylum 'Firmicutes. Int. J. Syst. Evol. Microbiol. 59, 1028–1035.
Käll, L., Krogh, A., Sonnhammer, E.L.L., 2007. Advantages of combined transmembrane topology and signal peptide prediction - the Phobius web server. Nucleic Acids Res. 35, W429–W432.
Kallnik, V., Bunescu, A., Sayer, C., Bräsen, C., Wohlgemuth, R., Littlechild, J., Siebers, B., 2014. Characterization of a phosphotriesterase-like lactonase from the hyperthermoacidophilic crenarchaeon *Vulcanisaeta moutnovskia*. J. Biotechnol. 190, 11–17.
Kim, H., Kwon, K.K., Seong, W., Lee, S.G., 2016. Multi-enzyme screening using a high-throughput genetic enzyme screening system. J. Vis. Exp. 114, e54059.
Kotik, M., Archelas, A., Wohlgemuth, R., 2012. Epoxide hydrolases and their application in organic synthesis. Curr. Org. Synth. 16, 451–482.
Kouker, G., Jaeger, K.E., 1987. Specific and sensitive plate assay for bacterial lipases. Appl. Environ. Microbiol. 53, 211–213.
Kovaleva, O.L., Merkel, A.Y., Novikov, A.A., Baslerov, R.V., Toshchakov, S.V., Bonch-Osmolovskaya, E.A., 2015. *Tepidisphaera mucosa* gen. Nov., sp. nov., a moderately thermophilic member of the class *Phycisphaerae* in the phylum *Planctomycetes*, and proposal of a new family, *Tepidisphaeraceae* fam. nov., and a new order, *Tepidisphaerales* ord. nov. Int. J. Syst. Evol. Microbiol. 65, 549–555.
Kublanov, I.V., Perevalova, A.A., Slobodkina, G.B., Lebedinsky, A.V., Bidzhieva, S.K., Kolganova, T.V., Kaliberda, E.N., Rumsh, L.D., Haertlé, T., Bonch-Osmolovskaya, E.A., 2009. Biodiversity of thermophilic prokaryotes with hydrolytic activities in hot springs of Uzon Caldera, Kamchatka (Russia). Appl. Environ. Microbiol. 75 (1), 286–291.
Lai, K.-K., Stogios, P.J., Vu, C., Xu, X., Cui, H., Molloy,S., et al. 2011. An inserted α/β subdomain shapes the catalytic pocket of *Lactobacillus johnsonii* cinnamoyl esterase. PLoS One 6:e23269.
Lakin, S.M., Dean, C., Noyes, N.R., Dettenwanger, A., Ross, A.S., Doster, E., Rovira, P., Abdo, Z., Jones, K.L., Ruiz, J., Belk, K.E., Morley, P.S., Boucher, C., 2017. MEGARes: an antimicrobial resistance database for high throughput sequencing. Nucleic Acids Res. 45, D574–D580.
Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M.R., Ahn, T.-H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S., Ussery, D.W., 2015. Insights from 20 years of bacterial genome sequencing. Funct. Integr. Genomics 15, 141–161.
Leis, B., Angelov, A., Mientus, M., Li, H.J., Pham, V.T.T., Lauinger, B., Bongen, P., Pietruszka, J., Gonçalves, L.G., Santos, H., Liebl, W., 2015. Identification of novel esterase-active enzymes from hot environments by use of the host bacterium *Thermus thermophilus*. Front. Microbiol. 6, 275.
Li, D., Liu, C.M., Luo, R., Sadakane, K., Lam, T., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct deBruijn graph. Bioinformatics 31, 1674–1676.
Liese, A., Hilterhaus, L., Kettling, U., Antranikian, G., 2016. Applied Biocatalysis: From Fundamental Science to Industrial Applications. Wiley-VCH, Weinheim, Germany.
Line, K., Isupov, M.N., Littlechild, J.A., 2004. The crystal structure of a (−) gamma-lactamase from an *Aureobacterium* species reveals a tetrahedral intermediate in the active site. J. Mol. Biol. 338, 519–532.
Liu, D., Evans, T., Zhang, F.Z., 2015. Applications and advances of metabolite biosensors for metabolic engineering. Metab. Eng. 31, 35–43.
Lobb, B., Kurtz, D.A., Moreno-Hagelsieb, G., Doxey, A.C., 2015. Remote homology and the functions of metagenomic dark matter. Front. Genet. 6, 234.
Mahr, R., Frunzke, J., 2016. Transcription factor-based biosensors in biotechnology: current state and future prospects. Appl. Microbiol. Biotechnol. 100, 79–90.
Mahr, R., von Boeselager, R.F., Wiechert, J., Frunzke, J., 2016. Screening of an *Escherichia coli* promoter library for a phenylalanine biosensor. Appl. Microbiol. Biotechnol. 100, 6739–6753.
Mandrich, L., Merone, L., Pezzullo, M., Cipolla, L., Nicotra, F., Rossi, M., Manco, G., 2005. Role of the N terminus in enzyme activity, stability and specificity in thermophilic esterases belonging to the HSL family. J. Mol. Biol. 345, 501–512.
Markowitz, V.M., Chen, I.M.A., Chu, K., Szeto, E., Palaniappan, K., Grechkin, Y., Ratner,

A., Jacob, B., Pati, A., Huntemann, M., Liolios, K., Pagani, I., Anderson, I., Mavromatis, K., Ivanova, N.N., Kyrpides, N.C., 2012. IMG/M: the integrated metagenome data management and comparative analysis system. Nucleic Acids Res. 40, D123–D129.
Menzel, P., Gudbergsdóttir, S.R., Rike, A.G., Lin, L., Zhang, Q., Contursi, P., Moracci, M., Kristjansson, J.K., Bolduc, B., Gavrilov, S., Ravin, N., Mardanov, A., Bonch-Osmolovskaya, E., Young, M., Krogh, A., Peng, X., 2015. Comparative metagenomics of eight geographically remote terrestrial hot springs. Microb. Ecol. 70, 411–424.
Menzel, P., Ng, K.L., Krogh, A., 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat. Commun. 7, 11257.
Meyer, H.-P., Eichhorn, E., Hanlon, S., Lütz, S., Schürmann, M., Wohlgemuth, R., Coppolecchia, R., 2013. The use of enzymes in organic synthesis and the life sciences: perspectives from the Swiss Industrial Biocatalysis Consortium (SIBC). Catal. Sci. Technol. 3, 29–40.
Michener, J.K., Thodey, K., Liang, J.C., Smolke, C.D., 2012. Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. Metab. Eng. 14, 212–222.
Namiki, T., Hachiya, T., Tanaka, H., Sakakibara, Y., 2012. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Res. 40, e155.
Nielsen, H.B., Almeida, M., Juncker, A.S., Rasmussen, S., Li, J., Sunagawa, S., Plichta, D.R., Gautier, L., Pedersen, A.G., Le Chatelier, E., Pelletier, E., Bonde, I., Nielsen, T., Manichanh, C., Arumugam, M., Batto, J.M., Quintanilha Dos Santos, M.B., Blom, N., Borruel, N., Burgdorf, K.S., Boumezbeur, F., Casellas, F., Doré, J., Dworzynski, P., Guarner, F., Hansen, T., Hildebrand, F., Kaas, R.S., Kennedy, S., Kristiansen, K., Kultima, J.R., Léonard, P., Levenez, F., Lund, O., Moumen, B., Le Paslier, D., Pons, N., Pedersen, O., Prifti, E., Qin, J., Raes, J., Sørensen, S., Tap, J., Tims, S., Ussery, D.W., Yamada, T., Renault, P., Sicheritz-Ponten, T., Bork, P., Wang, J., Brunak, S., Ehrlich, S.D., 2014. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat. Biotechnol. 32, 822–828.
Noguchi, H., Taniguchi, T., Itoh, T., 2008. Meta gene annotator: Detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res. 15, 387–396.
Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. MetaSPAdes: a new versatile metagenomic assembler. Genome Res. 27, 824–834.
Pevzner, P.A., Tang, H., Waterman, M.S., 2001. An Eulerian path approach to DNA fragment assembly. Proc. Natl. Acad. Sci. U. S. A. 98, 9748–9753.
Podosokorskaya, O.A., Bonch-Osmolovskaya, E.A., Godfroy, A., Gavrilov, S.N., Beskorovaynaya, D.A., Sokolova, T.G., Kolganova, T.V., Toshchakov, S.V., Kublanov, I.V., 2014. *Thermosipho activus* sp. nov., a thermophilic, anaerobic, hydrolytic bacterium isolated from a deep-sea sample. Int. J. Syst. Evol. Microbiol. 64, 3307–3313.
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., Le Paslier, D., Yakimov, M.M., Joachimiak, A., Ferrer, M., Golyshina, O.V., Savchenko, A., Golyshin, P.N., Yakunin, A.F., 2017. Activity screening of environmental metagenomic libraries reveals novel carboxylesterase families. Sci. Rep. 7, 44103.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, 590–596.
Raman, S., Rogers, J.K., Taylor, N.D., Church, G.M., 2014. Evolution-guided optimization of biosynthetic pathways. Proc. Natl. Acad. Sci. U. S. A. 111, 17803–17808.
Reed, B., Blazeck, J., Alper, H., 2012. Evolution of an alkane-inducible biosensor for increased responsiveness to short-chain alkanes. J. Biotechnol. 158, 75–79.
Rinaldi, S., Gori, A., Annovazzi, C., Ferrandi, E.E., Monti, D., Colombo, G., 2017. Unraveling energy and dynamics determinants to interpret protein functional plasticity: the limonene-1,2-epoxide-hydrolase case study. J. Chem. Inf. Model. 57, 717–725.
Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A., Dodsworth, J.A., Hedlund, B.P., Tsiamis, G., Sievert, S.M., Liu, W.T., Eisen, J.A., Hallam, S.J., Kyrpides, N.C., Stepanauskas, R., Rubin, E.M., Hugenholtz, P., Woyke, T., 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature 499, 431–437.
Rogers, J.K., Church, G.M., 2016. Genetically encoded sensors enable real-time observation of metabolite production. Proc. Natl. Acad. Sci. U. S. A. 113, 2388–2393.
Rohini Kumar, M., Saravanan, V.S., 2010. Candidate OP Phyla: Importance, Ecology and Cultivation prospects. Indian J. Microbiol. 50 (4), 474–477.
Sangwan, N., Xia, F., Gilbert, J.A., 2016. Recovering complete and draft population genomes from metagenome datasets. Microbiome 4, 1–11.
Sayer, C., Isupov, M.N., Bonch-Osmolovskaya, E., Littlechild, J.A., 2015a. Structural studies of a thermophilic esterase from a new Planctomycetes species, *Thermogutta terrifontis*. FEBS J. 282, 2846–2857.
Sayer, C., Szabo, Z., Isupov, M.N., Ingham, C., Littlechild, J.A., 2015b. The structure of a novel thermophilic esterase from the planctomycetes species, *Thermogutta terrifontis* reveals an open active site due to a minimal "cap" domain. Front. Microbiol. 6, 1–12.
Sayer, C., Finnigan, W., Isupov, M.N., Levisson, M., Kengen, S.W.M., Van Der Oost, J., Harmer, N.J., Littlechild, J.A., 2016. Structural and biochemical characterisation of *Archaeoglobus fulgidus* esterase reveals a bound CoA molecule in the vicinity of the active site. Sci. Rep. 6, 1–13.
Schnorr, K.M., Nygaard, P., Laloue, M., 1994. Molecular characterization of *Arabidopsis thaliana* cDNAs encoding three purine biosynthetic enzymes. Plant J. 6, 113–121.
Schnorr, K.M., Gaillard, C., Biget, E., Nygaard, P., Laloue, M., 1996a. A second form of adenine phosphoribosyltransferase in *Arabidopsis thaliana* with relative specificity towards cytokinins. Plant J. 9, 891–898.
Schnorr, K.M., Laloue, M., Hirel, B., 1996b. Isolation of cDNAs encoding two purine biosynthetic enzymes of soybean and expression of the corresponding transcripts in roots and root nodules. Plant Mol. Biol. 32, 751–757.
Seemann, T., 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069.
Siddiqui, K.S., 2015. Some like it hot, some like it cold: Temperature dependent biotechnological applications and improvements in extremophilic enzymes. Biotechnol. Adv. 33, 1912–1922.
Slobodkina, G.B., Kovaleva, O.L., Miroshnichenko, M.L., Slobodkin, A.I., Kolganova, T.V., Novikov, A.A., van Heerden, E., Bonch-Osmolovskaya, E.A., 2015. *Thermogutta terrifontis* gen. nov., sp. nov. and *Thermogutta hypogea* sp. nov., thermophilic anaerobic representatives of the phylum *Planctomycetes*. Int. J. Syst. Evol. Microbiol. 65, 760–765.
Söding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33, 244–248.
Swainston, N., Baici, A., Bakker, B.M., Cornish-Bowden, A., Fitzpatrick, P.F., Halling, P., Leyh, T.S., O'Donovan, C., Raushel, F.M., Reschel, U., Rohwer, J.M., Schnell, S., Schomburg, D., Tipton, K.F., Tsai, M.-D., Westerhoff, H.V., Wittig, U., Wohlgemuth, R., Kettner, C., 2018. STRENDA DB: enabling the validation and sharing of enzyme kinetics data. FEBS J. 285, 2193–2204.
Takami, H., Noguchi, H., Takaki, Y., Uchiyama, I., Toyoda, A., Nishi, S., Chee, G.J., Arai, W., Nunoura, T., Itoh, T., Hattori, M., Takai, K., 2012. A deeply branching thermophilic bacterium with an ancient Acetyl-CoA pathway dominates a subsurface ecosystem. PLoS One 7, e30559.
Tang, S.-Y., Qian, S., Akinterinwa, O., Frei, C.S., Gredell, J.A., Cirino, P.C., 2013. Screening for enhanced triacetic acid lactone production by recombinant *Escherichia coli* expressing a designed triacetic acid lactone reporter. J. Am. Chem. Soc. 135, 10099–10103.
Teather, R.M., Wood, P.J., 1982. Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. Appl. Environ. Microbiol. 43, 777–780.
Tirawongsaroj, P., Sriprang, R., Harnpichamchai, P., Thongaram, T., Champreda, V., Tanapongpipat, S., Pootanakit, K., Eurwilaichitr, L., 2008. Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. J. Biotechnol. 133, 42–49.
Toshchakov, S.V., Kublanov, I.V., Messina, E., Yakimov, M.M., Golyshin, P.N., 2015. Genomic analysis of pure cultures and communities. In: McGenity, T.J., Timmis, K.N., Nogales, B. (Eds.), Hydrocarbon and Lipid Microbiology ProtocolsGGSAC. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 5–27.
Trimble, W.L., Keegan, K.P., D'Souza, M., Wilke, A., Wilkening, J., Gilbert, J., Meyer, F., 2012. Short-read reading-frame predictors are not created equal: Sequence error causes loss of signal. BMC Bioinformatics 13, 183.
Uchiyama, T., Miyazaki, K., 2010. Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. Appl. Environ. Microbiol. 76, 7029–7035.
van Rossum, T., Kengen, S.W.M., van der Oost, J., 2013. Reporter-based screening and selection of enzymes. FEBS J. 280, 2979–2996.
van Rossum, T., Muras, A., Baur, M.J.J., Creutzburg, S.C.A., van der Oost, J., Kengen, S.W.M., 2017. A growth- and bioluminescence-based bioreporter for the *in vivo* detection of novel biocatalysts. Microb. Biotechnol. 10, 625–641.
Vincenzetti, S., Cambi, A., Neuhard, J., Schnorr, K., Grelloni, M., Vita, A., 1999. Cloning, expression, and purification of cytidine deaminase from *Arabidopsis thaliana*. Protein Expr. Purif. 15, 8–15.
Wei, Y., Swenson, L., Castro, C., Derewenda, U., Minor, W., Arai, H., Aoki, J., Inoue, K., Servin-Gonzalez, L., Derewenda, Z.S., 1998. Structure of a microbial homologue of mammalian platelet-activating factor acetylhydrolases: *Streptomyces exfoliatus* lipase at 1.9 Å resolution. Structure 6, 511–519.
Wei, Y., Contreras, J.A., Sheffield, P., Osterlund, T., Derewenda, U., Kneusel, R.E., Matern, U., Holm, C., Derewenda, Z.S., 1999. Crystal structure of brefeldin a esterase, a bacterial homolog of the mammalian hormone-sensitive lipase. Nat. Struct. Biol. 6, 340–345.
Wohlgemuth, R., 2010. Biocatalysis - key to sustainable industrial chemistry. Curr. Opin. Biotechnol. 21, 713–724.
Wohlgemuth, R., 2015. Epoxide hydrolysis. In: Faber, K., Fessner, W.-D., Turner, N.J. (Eds.), Science of Synthesis, Biocatalysis in Organic Synthesis 2. Georg Thieme Verlag KG, Stuttgart, pp. 529–555.
Wohlgemuth, R., 2017. Biocatalytic Process Design and Reaction Engineering. Chem. Biochem. Eng. Q. 31, 131–138.
Wohlgemuth, R., 2018. Horizons of systems biocatalysis and renaissance of metabolite synthesis. Biotechnol. J. 13 (6), 1700620.
Yang, G., Withers, S.G., 2009. Ultrahigh-throughput FACS-based screening for directed enzyme evolution. Chembiochem 10, 2704–2715.
Yang, Y., Zhong, C., Yooseph, S., 2015. SFA-SPA: a suffix array based short peptide assembler for metagenomic data. Bioinformatics 31, 1833–1835.
Yin, L., Bernhardt, P., Morley, K.L., Jiang, Y., Cheeseman, J.D., Purpero, V., Schrag, J.D., Kazlauskas, R.J., 2010. Switching catalysis from hydrolysis to perhydrolysis in *Pseudomonas fluorescens* esterase. Biochemistry 49, 1931–1942.
Yok, N.G., Rosen, G.L., 2011. Combining gene prediction methods to improve metagenomic gene annotation. BMC Bioinformatics 12, 20.
Youvan, D.C., Coleman, W.J., Bylina, E.J., 2002. Evolving and screening enzymes for new activities on polymer substrates. In: American Chemical Society, Polymer Preprints. Division of Polymer Chemistry, pp. 50–56.
Zalkin, H., Dixon, J.E., 1992. De novo purine biosynthesis. Prog. Nucl. Acid Res. Mol. Biol. 4, 259–287.
Zarafeta, D., Moschidi, D., Ladoukakis, E., Gavrilov, S., Chrysina, E.D., Chatziioannou, A., Kublanov, I., Skretas, G., Kolisis, F.N., 2016a. Metagenomic mining for thermostable esterolytic enzymes uncovers a new family of bacterial esterases. Sci. Rep. 6, 38886.
Zarafeta, D., Szabo, Z., Moschidi, D., Phan, H., Chrysina, E.D., Peng, X., Ingham, C.J.,

Kolisis, F.N., Skretas, G., 2016b. EstDZ3: a new esterolytic enzyme exhibiting remarkable thermostability. Front. Microbiol. 7, 1–14.
Zarafeta, D., Kissas, D., Sayer, C., Gudbergsdottir, S.R., Ladoukakis, E., Isupov, M.N., Chatziioannou, A., Peng, X., Littlechild, J.A., Skretas, G., Kolisis, F.N., 2016c. Discovery and characterization of a thermostable and highly halotolerant GH5 cellulase from an icelandic hot spring isolate. PLoS One 11, 1–18.
Zerbino, D.R., Birney, E., 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.
Zhang, J., Jensen, M.K., Keasling, J.D., 2015. Development of biosensors and their application in metabolic engineering. Curr. Opin. Chem. Biol. 28, 1–8.
Zhao, Y., Tang, H., Ye, Y., 2012. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. Bioinformatics 28, 125–126.
Zhu, W., Lomsadze, A., Borodovsky, M., 2010. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res. 38, e132.