

Plastics degradation by hydrolytic enzymes: the Plastics-Active Enzymes Database - PAZy

Patrick Buchholz¹, Hongli Zhang², Pablo Perez-Garcia², Lena-Luisa Nover², Jennifer Chow², Wolfgang R Streit², and Juergen Pleiss¹

¹Universität Stuttgart Universitätsbibliothek

²Universität Hamburg Fachbereich Chemie

December 11, 2021

Abstract

Petroleum based plastics are durable and accumulate in all ecological niches. Knowledge on enzymatic degradation is sparse. Today, less than 50 verified plastics-active enzymes are known. First examples of enzymes acting on the polymers polyethylene terephthalate (PET) and polyurethane (PUR) have been reported together with a detailed biochemical and structural description. Further, very few polyamide (PA) oligomer active enzymes are known. In this paper, the current known enzymes acting on the synthetic polymers PET and PUR are briefly summarized, their published activity data were collected and integrated into a comprehensive open access database. The Plastics-Active Enzymes Database (PAZy) represents an inventory of known and experimentally verified plastics-active enzymes. Almost 3000 homologues of PET-active enzymes were identified by profile hidden Markov models. Over 2000 homologues of PUR-active enzymes were identified by BLAST. Based on multiple sequence alignments, conservation analysis identified the most conserved amino acids, and sequence motifs for PET- and PUR-active enzymes were derived.

Plastics degradation by hydrolytic enzymes: the Plastics-Active Enzymes Database - PAZy

Patrick C. F. Buchholz¹, Hongli Zhang², Pablo Perez-Garcia², Lena-Luisa Nover², Jennifer Chow², Wolfgang R. Streit^{2*}, Jürgen Pleiss^{1*}

¹Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Stuttgart, Germany

²Department of Microbiology and Biotechnology, University of Hamburg, Hamburg, Germany

*Corresponding authors:

Jürgen Pleiss

Institute of Biochemistry and Technical Biochemistry,

University of Stuttgart, Allmandring 31,

D-70569 Stuttgart, Germany

Tel. +49 -711-685-63191

E-Mail: Juergen.Pleiss@itb.uni-stuttgart.de

ORCID: 0000-0003-1045-8202

Wolfgang R. Streit

Department of Microbiology and Biotechnology

Universität Hamburg, Ohnhorststraße 18 D-22609 Hamburg, Germany

Tel. +49-40-42816-463

E-Mail: wolfgang.streit@uni-hamburg.de

ORCID: 0000-0001-7617-7396

Keywords: hidden Markov model; hydrolases; metagenome; polyethylene terephthalate degradation; polyurethane degradation; sequence motif

Abstract

Petroleum based plastics are durable and accumulate in all ecological niches. Knowledge on enzymatic degradation is sparse. Today, less than 50 verified plastics-active enzymes are known. First examples of enzymes acting on the polymers polyethylene terephthalate (PET) and polyurethane (PUR) have been reported together with a detailed biochemical and structural description. Further, very few polyamide (PA) oligomer active enzymes are known. In this paper, the current known enzymes acting on the synthetic polymers PET and PUR are briefly summarized, their published activity data were collected and integrated into a comprehensive open access database. The Plastics-Active Enzymes Database (PAZy) represents an inventory of known and experimentally verified plastics-active enzymes. Almost 3000 homologues of PET-active enzymes were identified by profile hidden Markov models. Over 2000 homologues of PUR-active enzymes were identified by BLAST. Based on multiple sequence alignments, conservation analysis identified the most conserved amino acids, and sequence motifs for PET- and PUR-active enzymes were derived.

Introduction

Today, we face the global challenge of plastics pollution in nearly all environments. The pollution has meanwhile reached levels that will ultimately have impact on our food chain and well-being within the next decades. A recent study implied that about 399000 tons of plastics are present in the oceans alone, of which 69000 tons are microplastics¹. Thus, urgent actions need to be implemented for removal of plastics from the environment and by reducing the steady input into the environment. Whereas it is perhaps more likely that large pieces can be removed mechanically from ocean surfaces or terrestrial sites, smaller particles (microplastics) will remain there unless microbial or chemical degradation (i.e., weathering) will occur²⁻⁴. Plastic waste is a valuable raw material, therefore recycling is a promising alternative to incineration, either as a basis of synthesis of polymers or as a carbon source for fermentation⁵.

Petroleum-based plastics are in general extremely stable and durable; hence it is widely accepted that plastics do not degrade well in nature⁶, nor can be directly used in fermentation. The degradation processes described so far are slow, and it was shown that a PET bottle remains up to 48 years in the ocean until it is decomposed by microbial degradation⁷. Within this setting, it is reasonable to speculate that prior to microbial and enzymatic degradation, mechanical treatment (waves, wind, friction) and photodegradation by UV light (especially for aromatic ring-containing polymers such as PET and PS) break down the debris into microplastics, thereby increasing the surface area, which mediates microbial degradation. For more details on microplastics-associated bacteria and fungi, we refer to excellent reviews of the field of plastics ecology⁸⁻¹⁰. However, the colonization of microplastics does not necessarily indicate that the polymer is degraded, because additives are in general more bioavailable than the polymers. Therefore, measuring weight loss as an indicator for degradation might result in a false interpretation of the data¹¹, and in the conclusion that we already have many plastics-active enzymes from different microbial sources. A detailed search in the PubMed database revealed that today roughly 1500 publications address the topic of plastics degradation. However, less than 50 described the isolation and biochemical characterization of plastics-active enzymes (**Tables 1, S1-S3**). Nevertheless, while this obvious challenge can be met by better analytical techniques, the by far greater risk for misinterpretation of data comes from the unfiltered and non-critical use of the predicted plastic degrading microorganisms and consortia by not verified bioinformatic tools and pipelines.

For instance one recent study developed Hidden Markov models for some plastic degrading enzymes and

predicted a global distribution even though no such enzymes have been biochemically characterized¹². Others have developed phylogenetic trees and global distribution patterns by simply using automated literature searches without critical analyses of the data¹³. These very recent studies in high-ranking journals are perhaps only the tip of the iceberg, but clearly demonstrate that there is an urgent need for standardized and verified enzyme databases in this rapidly developing field. The non-critical and unfiltered use of many of the potential plastic degrading gene sequences ultimately leads to incorrect conclusions on the availability of plastic degrading enzymes and their role in nature. These studies do not only mislead researches, they furthermore suggest to environmentalists, policy and law makers and even to the broader public audience that we would have solutions for the global plastics problem, which we however do not have. Within this framework, the proposed PAZy database will be a reliable and very useful tool giving an overview on truly functional enzymes.

Notably today, only for polyethylene terephthalate (PET), polyurethane (PUR), and polyamide (PA), a rather small number of degrading enzymes are known, but none for other major polymers such as PVC, PE, PP, and PS, and most of the PUR-based polymers. The known plastics-degrading enzymes are hydrolases, often annotated as lipases, esterases, cutinases, amidases, or proteases (E.C. 3.1.x). However, we have still a limited understanding of the mechanism of enzymatic degradation. It is not clear to which extent bacteria have evolved specific enzymes that bind to the polymers and cleave the bonds similar to the processes that occur when cellulose or other biopolymers are degraded. It is supposed that plastics-degrading enzymes are exoenzymes, and it can be speculated that plastics-binding domains or proteins might contribute to degradation, similar to the role of cellulose binding domains or expansins in the degradation of cellulosic materials.

To advance the research field, we have collected information of the currently known and verified plastics-active enzymes in the Plastics-Active Enzymes Database (PAZy). It will serve as a comprehensive resource for the identification of further novel plastics-active enzymes, pathways, or microorganisms for plastics removal in industry and the environment. It will further help to advance improved circular use of the different plastic types. Finally, PAZy will in general be a valuable repository and tool in this emerging field of plastics research.

Methods

Data selection

Protein sequences with available Uniprot identifiers and known activity against PET or PUR were downloaded from the Plastics Microbial Biodegradation Database¹⁴ (PMBD, <http://pmbd.genome-mining.cn/home/>) and NCBI GenBank. Based on available biochemical and/or structural data, in total 44 protein sequences were selected from the PMBD, with 34 and 10 Uniprot identifiers for PET and PUR activity, respectively (as of November 2021).

PETase homologues

We are aware that there are controversial discussions about the term PETase, but we prefer to define all PET-active enzymes as PETases. Sixteen protein sequences for enzymes with known activity against PET were clustered using CD-HIT (version 4.6.8-1) at a threshold of 90% sequence identity and a word length of 5 to derive a reduced set of twelve centroid sequences^{15,16}. These protein sequences were aligned in a structure-guided multiple sequence alignment by T-COFFEE (version 11.00.8cbe486-1)¹⁷. A profile hidden Markov model (HMM) was derived from this multiple sequence alignment by HMMER (version 3.1b2, <http://hmmer.org>). The profile HMM was trimmed by selecting alignment columns that corresponded to the region between amino acid positions 32 and 274 in the PETase from *Ideonella sakaiensis* (*Is* PETase, Uniprot identifier A0A0K8P6T7) to avoid ambiguities at the N- and C-termini (**Table S4** and **Figure S1**). The profile HMM and the underlying multiple sequence alignment can be downloaded from <https://doi.org/10.18419/darus-2055>. This PETase-profile HMM was used to search both the NCBI non-redundant (nr) protein database and the Protein Data Bank (PDB) for an update of the *Lipase Engineering Database* (LED, <https://led.biocatnet.de>), which was previously established as a collection of protein se-

quences from α/β -hydrolases^{18–20}. Hits for the PETase-profile HMM were selected from the HMMER results with a minimal score of 100, a minimal profile coverage of 95%, and a maximum ratio of bias/score of 10%.

HMMER was also used to identify the C-terminal region for the Type IX secretion system sorting domain, using the profile HMM TIGR04183, which was derived from a multiple sequence alignment of 889 protein sequences in the TIGRFAM database (<http://tigrfams.jcvi.org/cgi-bin/index.cgi>), with an E-value cut-off below 1.

PURase homologues

Four protein sequences for enzymes with known activity against PUR served as queries for BLAST (blastp, version 2.10.0+) against the NCBI non-redundant (nr) protein database and the PDB²¹. BLAST performance was improved by multithreading with GNU/Parallel (version 20170622-1)²². The BLAST results were filtered by an E-value threshold of 10^{-10} and a minimal coverage of 50% to further update the LED.

Conservation analyses

The PETase-profile HMM was applied for a standard numbering scheme, by aligning the 2930 sequences of PETase homologues from the LED against the respective profile HMM and subsequently assigning the position numbers from the *Is* PETase reference sequence as *standard position* numbers. For conservation analysis of PETase homologues, the frequency of amino acid residues or gaps was counted at each standard position.

For the conservation analysis of PURase homologues from LED superfamilies 11 and 13, two multiple sequence alignments were generated using Clustal Omega (version 1.2.4)²³, and the frequency of amino acid residues or gaps was counted at selected positions.

Protein sequence networks

Sets of representative protein sequences were formed by clustering with CD-HIT to reduce the sample size and thus computational effort for pairwise sequence alignments. Values of pairwise sequence identity or similarity were calculated by the Needleman-Wunsch algorithm available in EMBOSS (version 6.6.0) with default gap opening and gap extension penalties of 10 and 0.5, respectively, and the substitution matrix BLOSUM62^{24,25}.

Collections of protein sequences were represented as protein sequence networks that depicted sequences as nodes connected by edges (lines). The edges in a protein sequence network were weighted by values of pairwise sequence identity or similarity. A threshold of the respective edge weights was chosen to select a subset of edges for the network. Protein sequence networks were visualized in Cytoscape (version 3.8.2) with the prefuse-force directed layout algorithm, taking the edge weights into account²⁶: edges of higher sequence identity or similarity were depicted preferably in closer vicinity to each other. The Python NetworkX package (version 1.11) was used to store the metadata of protein sequence networks in GraphML format, available for download at <https://doi.org/10.18419/darus-2054>²⁷.

Results

Update of the Lipase Engineering Database

We focus only on validated enzymes acting on the synthetic polymers PET and PUR. Detailed biochemical data of catalytically active enzymes (**Tables 1, S1-S3**), analysed sequences and structures of homologous proteins are comprised in the Plastics-Active Enzymes Database and accessible at <https://www.pazy.eu>. Within the PAZy infrastructure, the Lipase Engineering Database (LED, <https://led.biocatnet.de>) serves as the database for protein sequences and structures from different superfamilies of α/β -hydrolases and their sequence annotations, since all currently known enzyme activities towards PET or PUR were reported for α/β -hydrolases.

The updated LED contains 283,672 sequence entries and 1590 PDB entries (an increase of 3034 and 33 entries compared to the previously published LED version from June 2019, respectively). For the update of

the LED, sequences that shared at least 50% similarity were assigned to the same *superfamily* (**Table S5**). Sequences that shared at least 60% similarity were assigned to a *homologous family*; otherwise, they were assigned to a separate group containing all “singleton” sequences. A new superfamily was introduced for PURase homologues of PudaA from *Delftia acidovorans* as outlined in more detail below.

Polyethylene terephthalate (PET) active enzymes

Biochemical properties

Our literature searches identified a total of 34 enzymes that have been shown to catalyse the partial degradation of PET to oligomers or even to monomers, originating from four different bacterial phyla and one eukaryotic lineage (**Table 1**). No archaeal PETases have been functionally verified to date.

Many of the currently known PETases are thermostable enzymes, because the catalytic activity increases at temperatures close to the glass transition temperature (65 °C) of PET due to the formation of flexible and thus enzyme-accessible amorphous domains²⁸. Notably, few enzymes are active at lower temperatures implying they may play a role in cold-adapted PET degradation²⁹. However, all known native PETases have rather low catalytic activity toward PET.

Enzyme structures

All known PETases (EC number 3.1.1.101) are homologues of cutinases and are therefore annotated as di-lactone hydrolases, lipases, or esterases. In solution, PETases are supposed to be active as monomers³⁰. A total of twelve structures affiliated with different organisms are available in the PDB. For the best characterized examples LCC and *Is* PETase, multiple entries of variants have been made. All PETases consist of a single domain, the α/β -hydrolase fold, which is formed by a central twisted β -sheet, flanked by two layers of α -helices³¹, and thus belong to a class of small α/β -hydrolases that consist only of the core domain without a mobile lid²⁰. For a few PETase homologues from Bacteroidetes, an additional C-terminal sorting domain for the type IX secretion system has been annotated and was verified in the single structure published (**Figure S5, Table 1**). The type IX secretion system comprises several protein components³², and the corresponding C-terminal domain was also found in other polymer-active enzymes such as cellulases and endo-1,4- β -glucanases^{33–36}. PETases share a conserved catalytic triad of serine, histidine, and aspartic acid, and a GX-type oxyanion hole, which stabilizes the reaction intermediate³⁷. In the PETase homologues, the first oxyanion hole residue X is mostly a conserved aromatic residue such as tyrosine or phenylalanine. The second oxyanion-stabilizing residue is a conserved methionine following the serine of the catalytic triad. For the PETase from *I. sakaiensis*, several residues were suggested for substrate binding, such as an aromatic clamp formed by the first residue of the oxyanion hole and a second aromatic residue³⁸. In addition to this subsite I, a second subsite II was proposed from the interaction observed in a modelled complex with a PET monomer³⁹. Variants with increased catalytic activity were designed and tested. The most active enzymes are LCC variants, where the addition of disulfide bonds increased thermostability⁴⁰. The two LCC quadruple variants F243[WI]-D238C-S283C-Y127G, which include the additional disulfide bond, are among the most active PETases described to date.

Sequence network

For the comparison of PETase sequences, the profile HMM for PETases was used to identify the PETase core domain, and the sequences of all core domains were aligned without considering additional regions at the N- or C-termini (signal peptides or transport domains, respectively). In superfamily 1 of the LED, 31,560 sequences were annotated as GX-type, but only 2930 sequences were identified as PETase homologous by a profile HMM. At a threshold of 55% sequence similarity, the bacterial PETase core domains formed a large cluster, mainly originating from Actinobacteria or Proteobacteria (**Figure 1**). Most of the sequences from the PMBD were found in this cluster (**Figure S3**). In addition, a connected subgroup of PETase core domains from other bacterial phyla emerged, such as the PETase proteins from Bacteroidetes or Planctomycetes. Some homologues of PETase core domains occurred also in enzymes from extremophiles (**Figure S4**). The fungal PETase core domains such as the PETase homologues from *Fusarium* were separated from the bacterial

PETase core domains. At a higher threshold of 60% sequence similarity, the sequences for PETase core domains from Bacteroidetes or Planctomycetes emerged as a separated cluster (**Figure S5**).

Sequence motifs

The PETase-profile HMM was applied to analyse the conservation of amino acid residues in the 2930 PETase core domains annotated in the LED (**Table S6**) in comparison to the equivalent positions in the PETase from *Ideonella sakaiensis* (*Is* PETase, Uniprot identifier A0A0K8P6T7) and LCC (Uniprot identifier G9BY57). The catalytic triad, the previously suggested PET binding subsite I, which includes an aromatic clamp for possible substrate interaction, and PET binding subsite II from³⁹ were found to be highly conserved (**Table 2**). The extension of the second α -helix and the extended loop region, which were described previously as functionally relevant in *Is* PETase, were also found in several PETase homologues in the LED.

Using the position numbers from *Is* PETase, we suggest a typical PETase sequence motif written as follows (with X indicating an arbitrary amino acid): [YF]87, Q119, X₃ 139-141, S160, M161, W185, D206, H237, X₆ 242-247, followed by one of the previously published amino acid substitutions from⁴⁰. Interestingly, two sequences from an uncultured bacterium (NCBI: ACC95208.1) and *Alkalilimnicola ehrlichii* (NCBI: WP_116302080.1) were found to comprise the PETase sequence motif and W238, which was mentioned as an amino acid substitution for improved activity and substrate binding, and four additional sequences (from *Caldimonas manganoxidans*, NCBI: WP_019560450.1, from *C. taiwanensis*, NCBI:WP_062195544.1, from *Rhizobacter gummiphilus*, NCBI:WP_085749610.1, and from *Aquabacterium* sp., NCBI: MBI3384080.1) were found to comprise the PETase sequence motif and M241, which was mentioned as an amino acid substitution for improved thermostability. These six different and novel protein sequences, each selected by a sequence motif of seventeen amino acid positions in total, are proposed for upcoming studies on PETase activity.

Polyurethanes (PUR) active enzymes

Biochemical properties

Polyurethanes comprise numerous possible polymers of diverse composition, such as combinations of different isocyanates with different polyethers, polyesters, or polycarbonates⁴¹. The best studied PURases are α/β -hydrolases, as reviewed in more detail in⁴¹. Only ten characterized PUR-degrading enzymes (PURases) have been reported, yet. Four recently identified enzymes (LCC, TfCut-2, *Tcur_1278* T, and *Tcur0390*) are cutinases from Actinomyces, which are also active on PET and have a broad substrate profile⁴². Further bacterial lipases from Betaproteobacteria have been identified and characterized, such as *PueA* and *PueB* from *Pseudomonas chlororaphis*. Whereas all the above-mentioned studies identified lipases or esterases, earlier studies reported that commercially available peptidases and proteases might also degrade thin PUR films⁴³.

Enzyme structure

The ten known PURases belong to two groups (**Table S1**): four belong to the cutinases (LCC, TfCut-2, *Tcur_1278* T, and *Tcur0390*) and are similar to PETases. No crystal structures are available for the PURases *PueA* and *PueB* from *Pseudomonas chlororaphis*, but structures of homologues indicate that they belong to superfamily 11 of the LED, which in addition to the core domain has a mobile N-terminal lid, which might mediate binding to the substrate interface and substrate access, and an additional C-terminal β -sandwich domain. The four PETases and *PueA* and *PueB* from *P. chlororaphis* are GX types³⁷. Recently, a modelling study on the PURases from *Pseudomonas chlororaphis* predicted putative substrate binding sites for PUR-like substrates⁴⁴. However, a rearrangement of the substrate was observed upon the molecular simulation of the complex, which is an additional challenge for the identification of the substrate binding site⁴⁵. Interestingly, most of the substrate binding residues predicted for *PueA*⁴⁴ are conserved (**Table S7**).

The PURase from *Delftia acidovorans* (Uniprot identifier Q9WX47) does not belong to the GX type hydrolases, but has a sequence similarity to carboxylesterases of superfamily 13 and to the family PF00135 in Pfam, and thus belongs to the GGGX-type hydrolases³⁷. Other carboxylesterases in the LED are members of superfamily 4, which have a mobile lid between β -strand -4 and -3 of the core domain. Because the PURase

from *Delftia acidovorans* shared less than 50% sequence similarity to the sequences in the LED and due to the lack of experimental structure information, it was assigned to a new superfamily.

Sequence network

A threshold of 60% similarity was used to construct protein sequence networks for LED superfamilies 11 and 13 whose members originated from Proteobacteria only (**Figure 2**). Two disconnected sequence networks emerged: a network of 127 representative sequences of superfamily 11, which contains the GX-type PURases *PueA* and *PueB* from *Pseudomonas chlororaphis*, and a network of fifteen representative sequences of superfamily 13, which contains the GGGX-type the PURase *PudA* from *Delftia acidovorans*. In both superfamilies 11 and 13, the sequences originated mainly from Gammaproteobacteria, with the genus *Pseudomonas* being most frequently annotated in superfamily 11.

Sequence motifs

Two sequence motifs were reported previously in ⁴⁶ for the PURase from *Pseudomonas chlororaphis*, *PueA* (**Table 3**). The occurrence of the serine hydrolase motif (GXSXG) and the secretion signal sequence motif (GGXGXDXXX) were confirmed for the vast majority of all 2054 sequence entries in a multiple sequence alignment for superfamily 11 of the LED. Most PURase homologues from superfamily 11 have a GHSLG motif flanking the catalytic serine and secretion motif GGKGN DYLE. For sequences from superfamily 11 in the LED, the catalytic triad is formed by serine, aspartate, and histidine. In addition, most sequences in superfamily 11 (2039 out of 2054) matched the profile HMM for a RTX calcium-binding nonapeptide repeat (PFAM PF00353), which supports the previous suggestion of a Type I secretion system for protein translocation ⁴⁷.

Prominent amino acid positions in superfamily 13, which comprises homologues of the PURase *PudA* from *Delftia acidovorans*, include the GXSXG serine hydrolase motif, a catalytic triad of serine, aspartate, and histidine, and a putative PUR binding region at *PudA* positions 347 to 395 ⁴⁸. Many of these positions were found to be conserved within the sequences of superfamily 13 (**Table 4**). Most PURase homologues from superfamily 13 have the motif GES AG flanking the catalytic serine, the motif VPX₃G[ST]X₂DE at the catalytic glutamate, and the motif AXH X₃[LI]XY flanking the catalytic histidine. In addition, several positions in the putative PUR binding region were found to be conserved, including mostly hydrophobic amino acids (**Table S8**).

Discussion

Sequence annotations of plastics-active enzymes

Enzymes of different EC classes have been proposed to contribute to degradation of PET or PUR. α/β -hydrolases annotated as cutinases, esterases, or lipases were described to catalyze the hydrolysis of ester bonds and were collected in the PMBD ¹⁴, and peroxidases and laccases have been reported to enhance degradation of PUR ⁴⁹. For most types of plastics, however, knowledge on enzymatic degradation is missing, although materials such as polypropylene (PP) are produced in large scales and contribute extensively to global plastics pollution. In this paper, we focus on PET- and PUR-degrading α/β -hydrolases due to the availability of sequence information and detailed experimental data from literature and public databases ¹⁴.

Many PETase homologues were found in the NCBI non-redundant protein database, due to the sequencing efforts of the cutinase-expressing bacterial phyla such as Actinobacteria and Proteobacteria. In contrast, sequences from fungal origin are unrepresented. The usage of metagenomics is expected to further broaden the scope of currently known PETases ⁵⁰. The PETase sequence motif suggested herein (**Table 2**) is based on current knowledge from literature, such as the occurrence of an additional flexible loop region and amino acids that seem relevant for interaction with PET. The seven suggested PETase candidates can be used as starting points for wet-lab experiments, such as protein design experiments for improved PETase activity or thermostability.

Although PETases and PURases share the α/β -hydrolase fold as catalytic domain, their structure and their

oxyanion hole types differ. All the 2930 PETase homologues belong to a large family (36936 sequences) of GX-types, ³⁷ which consist of the core domain only (superfamily 1 in the Lipase Engineering Database,²⁰), whereas PURases belong to either of three superfamilies: superfamily 1, superfamily 11, which consists of GX-types with two additional domains, an N-terminal mobile lid and a C-terminal β -sandwich domain (2054 sequences), or to superfamily 13, which are carboxylesterases of the GGGX-type (44 sequences). Ample structural information is available for PET-degrading α/β -hydrolases, especially for LCC and *Is* PETase, whereas the structures of PUR-degrading α/β -hydrolases have not been resolved, yet. The former has also inspired the design of improved PETase variants, as recently demonstrated for variants based on LCC ⁴⁰. Previously, conserved subsite I, subsite II, and an extended loop region were identified in *Is* PETase and used for a systematic comparison with its homologues³⁹, which was confirmed by our conservation analysis of 2930 protein sequences. The profile HMM for PETases and the derived standard numbering scheme is available at the PAZy. It will help in the identification and comparison of amino acid positions reported in literature and will facilitate the design of new PETase variants. For PURases, homology models predict substrate binding regions^{44,45}. Recently, a putative PURase was identified in the Proteobacterium *Serratia liquefaciens* ⁵¹. Similarly, most putative PURases in the PAZy are from Proteobacteria (mainly from *Pseudomonas chlororaphis* and *Delftia acidovorans*).

Major challenges in searches for plastics-active enzymes

One of the major challenges will be the identification of novel enzymes for polymers for which none are currently known ^{11,52}. Thus, we urgently need enzymes acting on PE, PP, PVC, but also on polymeric PA and PU ether bonds. Further we realized the lack of a common PET or PUR model substrate, which would allow the direct comparison of the kinetic parameters of different plastics-active enzymes. In contrast, kinetic analysis is generally performed for typical esterase substrates such as pNP-caproate. This data, however, does not allow a reliable prediction of the actual plastics activity. The few enzymes that have been characterized using polymers were tested on different polymer types, and pre-treatment was used to enable better degradation (see **Table 1** for references). In addition, all kinetic data was recorded using single point measurements, thus the hydrolysis of the polymer could not be separated from the attachment of the enzyme to the polymer surface or from the hydrolysis of the resulting oligomers. Within this framework, a characterization of surface area and a control of surface properties such as crystallinity would be favourable to obtain better and more reliable kinetic data on the actual polymer.

The accumulation of verified plastics-active enzymes in databases with a reliable structure-function analysis will allow more predictive searches to rapidly and reliably identify novel and more active enzymes. Thereby it will allow to foster the search and development of novel pathways to create designer bugs using to solve the plastics problem.

How can databases contribute to a solution of the plastics waste problem?

For an efficient enzymatic degradation of plastics, we see four challenges. First, enzyme families other than α/β -hydrolases should be considered. For instance, laccases or peroxidases can also act on PUR⁴¹. First reports have been published but fell short in the identification of proteins and genes. Enzymatic or non-enzymatic degradation of other plastics components such as dyes or additives from commercial sources might need further investigations, too. Second, there is an increasing need for comparable data on plastics degradation. The comparability and reproducibility of data on enzymatic plastics degradation is impeded by the variety of possible substrates, e.g., in case of plastics built from several types of monomers. Furthermore, physical properties of plastic materials can differ remarkably among different commercial suppliers, e.g., thickness of plastic foils, number of additives, or crystallinity. Finally, incorrect annotation of genome and metagenome datasets has resulted in the accumulation of many false positive plastic degrading enzymes in various publications but also in PMBD. Removing these from the data bases is a major task.

Within this setting the PAZy database provides information on several well studied enzymes supplemented by the protein sequences and structures of homologous sequences, which enables the search for novel candidates and the design of enzyme variants. Most protein sequences and functional data are available for PETases,

for which several positions were already assessed experimentally for substrate binding or thermostability in earlier studies from literature (**Table 2**). The standard numbering scheme of PETase homologues facilitates the comparison of different amino acid positions from literature and the identification of sequence motifs for PETases, as shown for the comparison between *Is* PETase, LCC, and other PETase homologues. In upcoming studies, the PAZy database will be updated to cover sequences from other protein family databases than the LED, depending on the availability of structures and biochemical data. Finally, the underlying web platform of the PAZy database makes it easier to share knowledge on various plastics degrading enzymes in a more comparable way. Thus, it can be expected that the suggested sequence motifs for PETases or PURases will be refurbished, as more experimental data on residues for substrate binding or thermostability are made available in the future.

Acknowledgments

This work was in part supported by the BMBF within the programs MarBiotech (031B0562A), MetagenLig (FKZ 031B0571A and 031B0571B), LipoBiocat (031B0837B), and PlastiSea (031B867B).

References

1. Hohn S, Acevedo-Trejos E, Abrams JF, Fulgencio de Moura J, Spranz R, Merico A. The long-term legacy of plastic mass production. *Sci Total Environ* . 2020;746. doi:10.1016/j.scitotenv.2020.141115
2. Shaw DG, Day RH. Colour- and form-dependent loss of plastic micro-debris from the North Pacific Ocean. *Mar Pollut Bull* . 1994;28(1):39-43. doi:10.1016/0025-326X(94)90184-8
3. Mohammadian M, Allen NS, Edge M, Jones K. Environmental Degradation of Poly (ethylene Terephthalate). *Text Res J* . 1991;61(11):690-696. doi:10.1177/004051759106101109
4. Day M, Wiles DM. Photochemical degradation of poly(ethylene terephthalate). II. Effect of wavelength and environment on the decomposition process. *J Appl Polym Sci* . 1972;16(1):191-202. doi:10.1002/app.1972.070160117
5. Sadler JC, Wallace S. Microbial synthesis of vanillin from waste poly(ethylene terephthalate). *Green Chem* . 2021;23(13):4665-4672. doi:10.1039/d1gc00931a
6. Webb HK, Arnott J, Crawford RJ, Ivanova EP. Plastic degradation and its environmental implications with special reference to poly(ethylene terephthalate). *Polymers (Basel)* . 2013;5(1):1-18. doi:10.3390/polym5010001
7. Müller RJ, Kleeberg I, Deckwer WD. Biodegradation of polyesters containing aromatic constituents. *J Biotechnol* . 2001;86(2):87-95. doi:10.1016/S0168-1656(00)00407-7
8. Kirstein IV, Wichels A, Gullans E, Krohne G, Gerdts G. The plastisphere – Uncovering tightly attached plastic “specific” microorganisms. *PLoS One* . 2019;14(4). doi:10.1371/journal.pone.0215859
9. Yang Y, Liu W, Zhang Z, Grossart HP, Gadd GM. Microplastics provide new microbial niches in aquatic environments. *Appl Microbiol Biotechnol* . 2020;104(15):6501-6511. doi:10.1007/s00253-020-10704-x
10. Amaral-Zettler LA, Zettler ER, Mincer TJ. Ecology of the plastisphere. *Nat Rev Microbiol* . 2020;18(3):139-151. doi:10.1038/s41579-019-0308-0
11. Danso D, Chow J, Streit WR. Plastics: Environmental and biotechnological perspectives on microbial degradation. *Appl Environ Microbiol* . 2019;85(19):e01095-19. doi:10.1128/AEM.01095-19
12. Zrimec J, Kokina M, Jonasson S, Zorrilla F, Zelezniak A. Plastic-Degrading Potential across the Global Microbiome Correlates with Recent Pollution Trends. Kelly L, Newman DK, eds. *MBio* . 2021;12(5). doi:10.1128/MBIO.02155-21
13. Gambarini V, Pantos O, Kingsbury JM, Weaver L, Handley KM, Lear G. Phylogenetic Distribution of Plastic-Degrading Microorganisms. *mSystems* . 2021;6(1). doi:10.1128/MSYSTEMS.01112-20

14. Gan Z, Zhang H. PMBD: a Comprehensive Plastics Microbial Biodegradation Database. *Database (Oxford)* . 2019;2019. doi:10.1093/database/baz119
15. Li W, Godzik A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* . 2006;22:1658-1659. doi:10.1093/bioinformatics/btl158
16. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-Hit: accelerated for clustering the next-generation sequencing data. *Bioinformatics* . 2012;28(23):3150-3152. doi:10.1093/bioinformatics/bts565
17. Notredame C, Higgins DG, Heringa J. T-coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* . 2000;302(1):205-217. doi:10.1006/jmbi.2000.4042
18. Agarwala R, Barrett T, Beck J, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* . 2018;46(D1):D8-D13. doi:10.1093/nar/gkx1095
19. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res* . 2000;28(1):235-242. doi:10.1093/nar/28.1.235
20. Bauer TL, Buchholz PCF, Pleiss J. The modular structure of α - β -hydrolases. *FEBS J* . 2019;287(5):1035-1053. doi:10.1111/febs.15071
21. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* . 1990;215(3):403-410. doi:10.1016/S0022-2836(05)80360-2
22. Tange O. GNU parallel: the command-line power tool. *login USENIX Mag* . 2011;36(1):42-47.
23. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* . 2011;7(1):539. doi:10.1038/msb.2011.75
24. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* . 1970;48(3):443-453.
25. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet* . 2000;16(1):276-277. doi:10.1016/j.cocis.2008.07.002
26. Shannon P, Markiel A, Ozier O, et al. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res* . 2003;13(11):2498-2504. doi:10.1101/gr.1239303
27. Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using NetworkX. In: Varoquaux G, Vaught T, Millman J, eds. *Proceedings of the 7th Python in Science Conference* . ; 2008:11-15.
28. Alves NM, Mano JF, Balaguer E, Meseguer Dueñas JM, Gómez Ribelles JL. Glass transition and structural relaxation in semi-crystalline poly(ethylene terephthalate): A DSC study. *Polymer (Guildf)* . 2002;43(15):4111-4122. doi:10.1016/S0032-3861(02)00236-7
29. Zhang H, Dierkes R, Pérez-García P, et al. The abundance of mRNA transcripts of bacteroidetal polyethylene terephthalate (PET) esterase genes may indicate a role in marine plastic degradation. Published online August 11, 2021. doi:10.21203/RS.3.RS-567691/V2
30. Kitadokoro K, Thumarat U, Nakamura R, et al. Crystal structure of cutinase Est119 from *Thermobifida alba* AHK119 that can degrade modified polyethylene terephthalate at 1.76 Å resolution. *Polym Degrad Stab* . 2012;97(5):771-775. doi:10.1016/j.polymdegradstab.2012.02.003
31. David L, Cheah E, Cygler M, et al. The α/β hydrolase fold. *Protein Eng Des Sel* . 1992;5(3):197-211. doi:10.1093/protein/5.3.197
32. Lasica AM, Ksiazek M, Madej M, Potempa J. The Type IX Secretion System (T9SS): Highlights and Recent Insights into Its Structure and Function. *Front Cell Infect Microbiol* . 2017;7(MAY). doi:10.3389/fcimb.2017.00215

33. Raut MP, Couto N, Karunakaran E, Biggs CA, Wright PC. Deciphering the unique cellulose degradation mechanism of the ruminal bacterium *Fibrobacter succinogenes* S85. *Sci Rep* . 2019;9(1). doi:10.1038/s41598-019-52675-8
34. Suen G, Weimer PJ, Stevenson DM, et al. The complete genome sequence of fibrobacter succinogenes s85 reveals a cellulolytic and metabolic specialist. *PLoS One* . 2011;6(4). doi:10.1371/journal.pone.0018814
35. Arntzen M, Várnai A, Mackie RI, Eijsink VGH, Pope PB. Outer membrane vesicles from *Fibrobacter succinogenes* S85 contain an array of carbohydrate-active enzymes with versatile polysaccharide-degrading capacity. *Environ Microbiol* . 2017;19(7):2701-2714. doi:10.1111/1462-2920.13770
36. Brumm P, Mead D, Boyum J, et al. Functional annotation of *Fibrobacter succinogenes* S85 carbohydrate active enzymes. *Appl Biochem Biotechnol* . 2011;163(5):649-657. doi:10.1007/s12010-010-9070-5
37. Pleiss J, Fischer M, Peiker M, Thiele C, Schmid RD. Lipase engineering database: Understanding and exploiting sequence-structure-function relationships. *J Mol Catal - B Enzym* . 2000;10(5):491-508. doi:10.1016/S1381-1177(00)00092-8
38. Fecker T, Galaz-Davison P, Engelberger F, et al. Active Site Flexibility as a Hallmark for Efficient PET Degradation by *I. sakaiensis* PETase. *Biophys J* . 2018;114(6):1302-1312. doi:10.1016/j.bpj.2018.02.005
39. Joo S, Cho IJ, Seo H, et al. Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation. *Nat Commun* . 2018;9(1). doi:10.1038/s41467-018-02881-1
40. Tournier V, Topham CM, Gilles A, et al. An engineered PET depolymerase to break down and recycle plastic bottles. *Nature* . 2020;580(7802):216-219. doi:10.1038/s41586-020-2149-4
41. Magnin A, Pollet E, Phalip V, Avérous L. Evaluation of biological degradation of polyurethanes. *Biotechnol Adv* . 2020;39. doi:10.1016/j.biotechadv.2019.107457
42. Schmidt J, Wei R, Oeser T, et al. Degradation of Polyester Polyurethane by Bacterial Polyester Hydrolases. *Polymers (Basel)* . 2017;9(2):65. doi:10.3390/POLYM9020065
43. Phua SK, Castillo E, Anderson JM, Hiltner A. Biodegradation of a polyurethane in vitro. *J Biomed Mater Res* . 1987;21(2):231-246. doi:10.1002/jbm.820210207
44. do Canto VP, Thompson CE, Netz PA. Polyurethanases: Three-dimensional structures and molecular dynamics simulations of enzymes that degrade polyurethane. *J Mol Graph Model* . 2019;89:82-98. doi:10.1016/j.jmgm.2019.03.001
45. do Canto VP, Thompson CE, Netz PA. Computational studies of polyurethanases from *Pseudomonas*. *J Mol Model* . 2021;27(2). doi:10.1007/s00894-021-04671-x
46. Howard GT, Crother B, Vicknair J. Cloning, nucleotide sequencing and characterization of a polyurethanase gene (pueB) from *Pseudomonas chlororaphis*. *Int Biodeterior Biodegrad* . 2001;47(3):141-149. doi:10.1016/S0964-8305(01)00042-7
47. Bumba L, Masin J, Macek P, et al. Calcium-Driven Folding of RTX Domain β -Rolls Ratchets Translocation of RTX Proteins through Type I Secretion Ducts. *Mol Cell* . 2016;62(1):47-62. doi:10.1016/j.molcel.2016.03.018
48. Nomura N, Shigeno-Akutsu Y, Nakajima-Kambe T, Nakahara T. Cloning and sequence analysis of a polyurethane esterase of *Comamonas acidovorans* TB-35. *J Ferment Bioeng* . 1998;86(4):339-345. doi:10.1016/S0922-338X(99)89001-1
49. Ignat L, Ignat M, Ciobanu C, Doroftei F, Popa VI. Effects of flax lignin addition on enzymatic oxidation of poly(ethylene adipate) urethanes. *Ind Crops Prod* . 2011;34(1):1017-1028. doi:10.1016/j.indcrop.2011.03.010

50. Danso D, Schmeisser C, Chow J, et al. New insights into the function and global distribution of polyethylene terephthalate (PET)-degrading bacteria and enzymes in marine and terrestrial metagenomes. *Appl Environ Microbiol* . 2018;84(8). doi:10.1128/AEM.02773-17
51. Salgado CA, Almeida FA de, Barros E, Baracat-Pereira MC, Baglinière F, Vanetti MCD. Identification and characterization of a polyurethanase with lipase activity from *Serratia liquefaciens* isolated from cold raw cow's milk. *Food Chem* . 2021;337. doi:10.1016/j.foodchem.2020.127954
52. Wei R, Zimmermann W. Microbial enzymes for the recycling of recalcitrant petroleum-based plastics: how far are we? *Microb Biotechnol* . 2017;10(6):1308-1322. doi:10.1111/1751-7915.12710
53. Yoshida S, Hiraga K, Takehana T, et al. A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science (80-)* . 2016;351(6278):1196-1199. doi:10.1126/science.aad6359
54. Han X, Liu W, Huang JW, et al. Structural insight into catalytic mechanism of PET hydrolase. *Nat Commun* . 2017;8(1). doi:10.1038/s41467-017-02255-z
55. Austin HP, Allen MD, Donohoe BS, et al. Characterization and engineering of a plastic-degrading aromatic polyesterase. *Proc Natl Acad Sci U S A* . 2018;115(19):E4350-E4357. doi:10.1073/pnas.1718804115
56. Wei R, Song C, Gräning D, et al. Conformational fitting of a flexible oligomeric substrate does not explain the enzymatic PET degradation. *Nat Commun 2019 101* . 2019;10(1):1-4. doi:10.1038/s41467-019-13492-9
57. Son HF, Cho IJ, Joo S, et al. Rational Protein Engineering of Thermo-Stable PETase from *Ideonella sakaiensis* for Highly Efficient PET Degradation. *ACS Catal* . 2019;9(4):3519-3526. doi:10.1021/ACSCATAL.9B00568
58. C L, C S, S Z, R W, CC Y. Structural and functional characterization of polyethylene terephthalate hydrolase from *Ideonella sakaiensis*. *Biochem Biophys Res Commun* . 2019;508(1):289-294. doi:10.1016/J.BBRC.2018.11.148
59. Haernvall K, Zitzenbacher S, Yamamoto M, Schick MB, Ribitsch D, Guebitz GM. A new arylesterase from *Pseudomonas pseudoalcaligenes* can hydrolyze ionic phthalic polyesters. *J Biotechnol* . 2017;257:70-77. doi:10.1016/j.jbiotec.2017.01.012
60. A B, S T, E K-G, et al. A Novel Polyester Hydrolase From the Marine Bacterium *Pseudomonas aestusnigri* - Structural and Functional Insights. *Front Microbiol* . 2020;11. doi:10.3389/FMICB.2020.00114
61. Ronkvist ÅM, Xie W, Lu W, Gross RA. Cutinase-Catalyzed Hydrolysis of Poly(ethylene terephthalate). *Macromolecules* . 2009;42(14):5128-5138. doi:10.1021/MA9005318
62. Sulaiman S, Yamato S, Kanaya E, et al. Isolation of a novel cutinase homolog with polyethylene terephthalate-degrading activity from leaf-branch compost by using a metagenomic approach. *Appl Environ Microbiol* . 2012;78(5):1556-1562. doi:10.1128/AEM.06725-11
63. S S, DJ Y, E K, Y K, S K. Crystal structure and thermodynamic and kinetic stability of metagenome-derived LC-cutinase. *Biochemistry* . 2014;53(11):1858-1869. doi:10.1021/BI401561P
64. Shirke AN, White C, Englaender JA, et al. Stabilizing Leaf and Branch Compost Cutinase (LCC) with Glycosylation: Mechanism and Effect on PET Hydrolysis. *Biochemistry* . 2018;57(7):1190-1200. doi:10.1021/acs.biochem.7b01189
65. Dresler K, Van Den Heuvel J, Müller RJ, Deckwer WD. Production of a recombinant polyester-cleaving hydrolase from *Thermobifida fusca* in *Escherichia coli*. *Bioprocess Biosyst Eng* . 2006;29(3):169-183. doi:10.1007/s00449-006-0069-9
66. Müller R-J, Schrader H, Profe J, Dresler K, Deckwer W-D. Enzymatic Degradation of Poly(ethylene terephthalate): Rapid Hydrolyse using a Hydrolase from *T. fusca*. *Macromol Rapid Commun* . 2005;26(17):1400-1405. doi:10.1002/MARC.200500410

67. Kleeberg I, Hetz C, Kroppenstedt RM, Müller RJ, Deckwer WD. Biodegradation of aliphatic-aromatic copolyesters by *Thermomonospora fusca* and other thermophilic compost isolates. *Appl Environ Microbiol* . 1998;64(5):1731-1735. doi:10.1128/aem.64.5.1731-1735.1998
68. Kleeberg I, Welzel K, VandenHeuvel J, Müller RJ, Deckwer WD. Characterization of a new extracellular hydrolase from *Thermobifida fusca* degrading aliphatic-aromatic copolyesters. *Biomacromolecules* . 2005;6(1):262-270. doi:10.1021/bm049582t
69. Acero EH, Ribitsch D, Steinkellner G, et al. Enzymatic Surface Hydrolysis of PET: Effect of Structural Diversity on Kinetic Properties of Cutinases from *Thermobifida*. *Macromolecules* . 2011;44(12):4632-4640. doi:10.1021/MA200949P
70. Chen S, Tong X, Woodard RW, Du G, Wu J, Chen J. Identification and characterization of bacterial cutinase. *J Biol Chem* . 2008;283(38):25854-25862. doi:10.1074/jbc.M800848200
71. Su L, Woodard RW, Chen J, Wu J. Extracellular location of *Thermobifida fusca* cutinase expressed in *Escherichia coli* BL21(DE3) without mediation of a signal peptide. *Appl Environ Microbiol* . 2013;79(14):4192-4198. doi:10.1128/AEM.00239-13
72. Lykidis A, Mavromatis K, Ivanova N, et al. Genome sequence and analysis of the soil cellulolytic actinomycete *Thermobifida fusca* YX. *J Bacteriol* . 2007;189(6):2477-2486. doi:10.1128/JB.01899-06
73. Furukawa M, Kawakami N, Tomizawa A, Miyamoto K. Efficient Degradation of Poly(ethylene terephthalate) with *Thermobifida fusca* Cutinase Exhibiting Improved Catalytic Activity Generated using Mutagenesis and Additive-based Approaches. *Sci Rep* . 2019;9(1). doi:10.1038/s41598-019-52379-z
74. Roth C, Wei R, Oeser T, et al. Structural and functional studies on a thermostable polyethylene terephthalate degrading hydrolase from *Thermobifida fusca*. *Appl Microbiol Biotechnol* . 2014;98(18):7815-7823. doi:10.1007/s00253-014-5672-0
75. Wei R, Oeser T, Schmidt J, et al. Engineered bacterial polyester hydrolases efficiently degrade polyethylene terephthalate due to relieved product inhibition. *Biotechnol Bioeng* . 2016;113(8):1658-1665. doi:10.1002/bit.25941
76. Huang YC, Chen GH, Chen YF, Chen WL, Yang CH. Heterologous expression of thermostable acetylxylan esterase gene from *Thermobifida fusca* and its synergistic action with xylanase for the production of xylooligosaccharides. *Biochem Biophys Res Commun* . 2010;400(4):718-723. doi:10.1016/j.bbrc.2010.08.136
77. Hegde K, Veeranki VD. Production optimization and characterization of recombinant cutinases from *thermobifida fusca* sp. NRRL B-8184. *Appl Biochem Biotechnol* . 2013;170(3):654-675. doi:10.1007/s12010-013-0219-x
78. Ribitsch D, Hromic A, Zitzenbacher S, et al. Small cause, large effect: Structural characterization of cutinases from *Thermobifida cellulolytica*. *Biotechnol Bioeng* . 2017;114(11):2481-2488. doi:10.1002/bit.26372
79. Kitadokoro K, Kakara M, Matsui S, et al. Structural insights into the unique polylactate-degrading mechanism of *Thermobifida alba* cutinase. *FEBS J* . 2019;286(11):2087-2098. doi:10.1111/febs.14781
80. Wei R, Oeser T, Then J, et al. Functional characterization and structural modeling of synthetic polyester-degrading hydrolases from *Thermomonospora curvata*. *AMB Express* . 2014;4(1):1-10. doi:10.1186/s13568-014-0044-9
81. Ribitsch D, Acero EH, Greimel K, et al. A New Esterase from *Thermobifida halotolerans* Hydrolyses Polyethylene Terephthalate (PET) and Polylactic Acid (PLA). *Polym 2012, Vol 4, Pages 617-629* . 2012;4(1):617-629. doi:10.3390/POLYM4010617
82. Hu X, Thumarat U, Zhang X, Tang M, Kawai F. Diversity of polyester-degrading bacteria in compost and molecular analysis of a thermoactive esterase from *Thermobifida alba* AHK119. *Appl Microbiol Biotechnol* . 2010;87(2):771-779. doi:10.1007/s00253-010-2555-x

83. Kawai F, Oda M, Tamashiro T, et al. A novel Ca²⁺-activated, thermostabilized polyesterase capable of hydrolyzing polyethylene terephthalate from *Saccharomonospora viridis* AHK190. *Appl Microbiol Biotechnol* . 2014;98(24):10053-10064. doi:10.1007/s00253-014-5860-y
84. Oda M, Yamagami Y, Inaba S, et al. Enzymatic hydrolysis of PET: functional roles of three Ca²⁺ ions bound to a cutinase-like enzyme, Cut190*, and its engineering for improved activity. *Appl Microbiol Biotechnol* . 2018;102(23):10067-10077. doi:10.1007/s00253-018-9374-x
85. Ribitsch D, Heumann S, Trotscha E, et al. Hydrolysis of polyethyleneterephthalate by p-nitrobenzylesterase from *Bacillus subtilis*. *Biotechnol Prog* . 2011;27(4):951-960. doi:10.1002/BTPR.610
86. Carniel A, Valoni É, Nicomedes J, Gomes A da C, Castro AM de. Lipase from *Candida antarctica* (CALB) and cutinase from *Humicola insolens* act synergistically for PET hydrolysis to terephthalic acid. *Process Biochem* . 2017;59:84-90. doi:10.1016/J.PROCBIO.2016.07.023

Tables

TABLE 1: Currently known* and active PET hydrolases together with recent publications of variants and structural or molecular analyses. All (The majority) of WT enzymes listed were verified for their PET hydrolytic activities by either PET films, PET powder PET bottles, PET coupons, PET nanoparticles or various synthesized model polyester polymers. Only isolates were included that are available in public strain repositories.

Microbial host, enzyme, gene ^d	Ref.	Genbank, Uniprot ID ^a	PDB structure	WT Enzymes and variants tested	Activity on PET foil, powder, bottle ^b
Proteobacteria-affiliated PET active enzymes					
<i>Ideonella sakaiensis</i> 201-F6, ISPETase, <i>ISF6_4831</i>	53	<i>A0A0K8P6T7</i>		WT	
	54		<i>5XFZ 5XFY</i> <i>5XG0 5XH2</i> <i>5XH3</i>	R103G/S131A, S185H, C174S, C210S, W156A, S131A, I179A, W130A, M132A, Y58A, T59A R132G/S160A	
	55		<i>6EQE 6EQD</i> <i>6EQG 6EQF</i> <i>6EQH</i>	W185A, S238F/W159H	+, crystallinity changes observed
	39, 56		<i>5XJH 5YNS</i>	R280A, C203A/C239A	+, only relative data using 200 nM of enzyme

Microbial host, enzyme, gene ^d	Ref.	Genbank, Uniprot ID ^a	PDB structure	WT Enzymes and variants tested	Activity on PET foil, powder, bottle ^b
	57		<i>6IJ4 6IJ3 6IJ6 6IJ5</i>	S160A, D206A, H237A, Y87A, M161A, W185A, I208A, W159A, S238A, N241A, R280A, W159H, S238F	14-fold increased activity of S121E/D186H/R280A over WT; 83 μM TA and 37μM MHET released after 1-10 days
	58		<i>6ILW</i>	N241F, W159F, S93M	n.d.
	38		<i>6ANE</i>	C176A, C212A, C176A-C212A	n.d.
PET4, <i>Marinobacter nanhaiticus</i> D15-8W, Gene: J057_15340	50	<i>N6VY44-9ALTE</i>		WT enzyme	+
PET5, <i>Oleispira antarctica</i> RB-8 LipA	50	<i>R4YKL9-OLEAN</i>		WT enzyme	+
PET6, <i>Vibrio gazogenes</i> Gene: BSQ33_03270	50	<i>A0A1Z2SIQ1-VIBGA</i>		WT enzyme	+
PET12, <i>Polyangium brachysporum</i> Gene: AAW51_2473	50	<i>A0A0G3BI90-9BURK</i>		WT	+
<i>P. pseudoalcaligenes</i> , ⁵⁹ DSM 50188 PpCutA		<i>KU695574</i>			n.d.
<i>P. pelagia</i> , DSM 25163 PpelaLip	59	<i>KU695573</i>			n.d.
<i>P. aestusnigri</i> VGXO14, B7O88_11480, PE-H	60	<i>A0A1H6AD45</i>	<i>6SBN</i>	WT, Y250S S171A, D217A, H249A, G254S, S256N, 257S, Y258N, N250S, G241P, Y250S, C241P	6 mg MHET/L from PET film (WT and N250S, G241P, Y250S, C241P) 500nM enzyme ⁻¹ ; and 0.1 mg/MHET from PET bottle for Y250S in 24h x 500nM enzyme ⁻¹ solubilized 250 μm thick films in 96 hours
<i>P. mendocina</i> ATCC 53552 PMC	61	<i>N20M5AZM016</i>	<i>2FX5</i>		

Microbial host, enzyme, gene ^d	Ref.	Genbank, Uniprot ID ^a	PDB structure	WT Enzymes and variants tested	Activity on PET foil, powder, bottle ^b
Actinobacterial enzymes					
LCC, leaf compost metagenome	62	<i>G9BY57</i> AEV21261.1			12 mg TA _{eq.} X h ⁻¹ x mg enzyme ⁻¹ with WT enzyme
	63		<i>4EB0</i>		
	64			N197Q, N266Q, N239G, LCC-G	
	40		6THS 6THT	S165A, ICCG-S165A F243I/D238C/S283C/ Y127G (ICCG), F243I/D238C/S283C/ (ICCM), F243W/D238C/ S283C/Y127G (WCCG) and F243W/D238C/S283C/N246M (WCCM),	105.6 ± 3.9 mg TA _{eq.} X h ⁻¹ x mg enzyme ⁻¹ , on commercial CNP46M with best variant
BhrPETase from HRB29 bacterium					
<i>T. fusca</i> DSM43793 Tfh, BTA-1	65, 66, 67, 68	<i>Q6A0I4.THEFU</i> <i>AJ810119</i>		WT	+
<i>T. fusca</i> DSM43793 Tfh, BTA-2	65, 66, 67, 68	<i>AJ810119</i>			+
<i>T. fusca</i> DSM 44342, Tfh42_Cut1	69	<i>E9LVIO.THEFU</i> <i>ADV92528.1</i>			41 mmol TA/mmol enzyme in 120 hours
<i>T. fusca</i> , (strain xy) WSH03-11, Tfu_0882	70, 71, 72	<i>Q47RJ7.-THEFY</i>			n.d.
<i>T. fusca</i> , (strain xy) WSH03-11, Tfu_0883	70	<i>Q47RJ6.-THEFY</i>			n.d.

Microbial host, enzyme, gene ^d	Ref.	Genbank, Uniprot ID ^a	PDB structure	WT Enzymes and variants tested	Activity on PET foil, powder, bottle ^b
<i>T. fusca</i> , TfCut-2 (Cut2-kw3)	73	<i>E5BBQ3- THEFU</i>		G62A/F209A	31 ± 0.1 nmol min ⁻¹ cm ⁻² on films; 12-fold better than WT; lcPET film (200 μm) 97 ± 1.8% in 30 h
	74		4CG1 4CG2 4CG3		
	75		4CG1	G62A/I213S, G62A	42% weight loss after 50 h on film G62A/I213S, G62A 2.7-fold better than WT n.d.
<i>T. fusca</i> NTU22, TfAXE	76	<i>ADM47605.1</i>			n.d.
<i>T. fusca</i> NRRL B-8184, Cut1	77	<i>JN129499.1</i>			n.d.
<i>T. fusca</i> NRRL B-8184, Cut2	77	<i>JN129500.1</i>			n.d.
<i>T. cellulosilytica</i> DSM44535, Thc_Cut1	78, 69	<i>ADV92526.1</i>	<i>5LUI</i>		56 mmol TA/mmol enzyme in 120 hours
<i>T. cellulosilytica</i> DSM44535, Thc_Cut2	78, 69	<i>ADV92527.1</i>	<i>5LUJ 5LUL 5LUK</i>	R29N/A30V, R19S/R29N/A30V	5 mmol TA/mmol enzyme in 120 hours
<i>T. alba</i> AHL119, Est119, est2	79	<i>F7IX06</i>	<i>6AID 3WYN 3VIS</i>		n.d.
<i>T. curvata</i> , DSM43183 <i>Tcur_1278</i>	80	<i>D1A9G5 ACY96861.1</i>			n.d.
<i>T. curvata</i> , DSM43183 <i>Tcur0390</i>	80	<i>ACY95991.1</i>			n.d.
<i>T. halotolerans</i> , <i>Thh_Est</i>	81	<i>H6WX58</i>			21.5 mmol TA released /mol enzyme; 2 hours
<i>T. alba</i> , <i>Est1</i> (<i>Hydrolase 4</i>)	82	<i>BAI99230</i>			n.d.

Microbial host, enzyme, gene ^d	Ref.	Genbank, Uniprot ID ^a	PDB structure	WT Enzymes and variants tested	Activity on PET foil, powder, bottle ^b
<i>Saccharomonospora</i> (Thermoactinomyces) <i>viridis</i> AHK190	83, 83, 84	W0TJ64 AB728484	4WFK 4WFI 4WFJ	S226P, S226P/R228S, Q138A/D250C- E296C/Q123H/N202H, E296, at N250 and	30% increase in Q138A/D250C- E296C/Q123H/N202H, variant, but no kinetic data
Firmicutes <i>Bacillus subtilis</i> 4P3-11, BsEstB	85	ADH43200.1			nd
Bacteroidetes <i>Aequorivita</i> sp. CIP111184, PET27	29	WP_111881932			Active on foil and powder
<i>Kaistella</i> (Chryseobacteriu) <i>jeonii</i> , PET30	29	WP_039353427	7PZJ		Active on foil and powder
Metagenome, probably bacterial and phylogenetically unassigned PET2, from metagenome, affiliated no obvious affiliation, lipIAF5-2	50	C3RYL0 ACC95208.1			+
PET active enzymes from Eukaryotes <i>Candida antarctica</i> lipase B, CalB	61	LIPB_PSEA2		WT	solubilized 250 µm thick films in 96 hours
<i>Fusarium solani</i> FsC	61	AAA33335.1	1OXM	WT WT	+ solubilized 250 µm thick films in 96 hours
<i>Thermomyces (Humicola) insolens</i> HiC	61	A0A075B5G4	4OYY, 4OYL	WT	solubilized 250 µm thick films in 96 hours
	86			WT	+

a) , active link to NCBI Pubmed database

b) n.d., not determined; - , no activity observed;+ , activity observed but not quantified; crystallinity of PET films, bottles and samples and experimental conditions varied largely in the different studies and thus makes direct comparison difficult.

TABLE 2 : A selection of amino acid positions in the 2930 PETase homologues from the LED. Amino acid residues that occurred in less than 50% of the sequences are listed in brackets for comparison with their previous mentions in literature. Amino acid substitutions “in LCC” refer to the position numbers used in ⁴⁰. Some positions only occurred in a small subset of all PETase homologues and are thus indicated as gap symbols (-). Percentages indicate values rounded to integers. Sixteen positions and their corresponding amino acid symbols marked in red indicate the suggested PETase sequence motif, whereas the ones marked in green indicate additional positions that were used to select the sequences mentioned in the main text.

Position in <i>Is</i> PETase	Amino acid or gap symbol	Annotations or substitutions	References
87	Y 51%, F 46%	Subsite I, aromatic clamp, oxyanion hole	38,39
88	T 76% (V 7%, L 5%, M 1%)	Subsite II	39
		Thermostability: T96 M in LCC	40
89	A 70% (G 12%, S 10%)	Subsite II	39
119	Q 69% (F8%, Y4%, W2%)	Subsite I	39
		Thermostability: Y127 G in LCC	40
139	Gap 91%	Extension in α -helix 2	38
140	Gap 80%	Extension in α -helix 2	38
141	Gap 90%	Extension in α -helix 2	38
159	H 87% (W9%)	Subsite II	38,39
160	S 100%	Catalytic triad	38
161	M 94%	Subsite I, oxyanion hole	39
185	W 77% (Y12%)	Subsite I, aromatic clamp, ”wobbly tryptophane”	38,39
206	D 100%	Catalytic triad	38
208	V 54% (I 36%)	Substrate interaction	38
237	H 100%	Catalytic triad	38
238	F 64% (L 11%, S 8%, Y 3%)	Subsite II	38,39
		Binding & activity: F243 I in LCC	40
		Binding & activity: F243 W in LCC	40
241	N 56% (Q 12%)	Subsite II	39
		Thermostability: N246 D in LCC	40
		Thermostability: N246 M in LCC	40

Position in <i>IsPETase</i>	Amino acid or gap symbol	Annotations or substitutions	References
242	(S 36%, T 32%, G 5%, I 4%)	Extended loop	38,39
243	(S 40, G 11%, T 8%)	Extended loop	39
244	N 74% (G 3%, Y 3%)	Extended loop	39
245	- 87%	Extended loop	39
246	- 89%	Extended loop	39
247	- 87%	Extended loop	39

TABLE 3 : Amino acid positions mentioned in ⁴⁶and their occurrence in 2054 PURase homologues from the LED (superfamily 11). Position numbers refer to PueA from *Pseudomonas chlororaphis* , i.e., Uniprot identifier A1Z374. Percentages indicate values rounded to integers.

Position in PueA	Amino acid	Annotation
204	G 100%	Serine hydrolase motif GXSXG
205	H 99%	Serine hydrolase motif GXSXG
206	S 100%	Serine hydrolase motif GXSXG , catalytic triad
207	L 99%	Serine hydrolase motif GXSXG
208	G 100%	Serine hydrolase motif GXSXG
254	D 100%	Catalytic triad
312	H 100%	Catalytic triad
381	G 99%	Secretion motif GGXGXDXXX
382	G 98%	Secretion motif GGXGXDXXX
383	K 39%, R 18%, A 16%, S 13%	Secretion motif GGXGXDXXX
384	G 98%	Secretion motif GGXGXDXXX
385	N 78%, A 16%	Secretion motif GGXGXDXXX
386	D 99%	Secretion motif GGXGXDXXX
387	Y 57%, F 42%	Secretion motif GGXGXDXXX
388	L 72%, I 27%	Secretion motif GGXGXDXXX
389	E 91%	Secretion motif GGXGXDXXX

TABLE 4 : Amino acid positions mentioned in ⁴⁸and adjacent amino acid positions that occur in at least 50% of the 44 PURase homologues from the LED (superfamily 13). Position numbers refer to Puda from *Delftia acidovorans* , i.e., Uniprot identifier Q9WX47. Similar amino acid residues occurring in less than 50% of the sequences are indicated in brackets for comparison. Positions of a putative substrate binding region are listed as supporting information for comparison (**Table S5**).

Position in Puda	Amino acid	Annotation
223	G 100%	Serine hydrolase motif GXSXG
224	E 90%	Serine hydrolase motif GXSXG
225	S 100%	Serine hydrolase motif GXSXG , catalytic triad
226	A 97%	Serine hydrolase motif GXSXG
227	G 100%	Serine hydrolase motif GXSXG
340	V 100%	
341	P 97%	
342	V 64%	
343	(I 33%, V 30%, M 26%)	

Position in Puda	Amino acid	Annotation
345	G 100%	
346	S 50% (T 42%)	
347	N 71%	
349	D 92%	
350	E 100%	Catalytic triad
457	A 95%	
458	A 66%	
459	H 100%	Catalytic triad
462	E 78%	
463	L 69% (I 30%)	
464	Q 83%	
465	Y 92%	
466	L 83%	

Figures

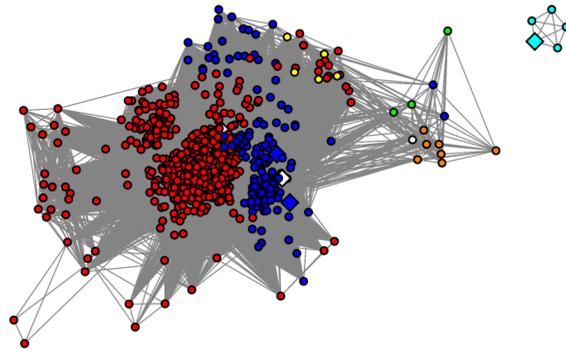


FIGURE 1. Network representation for 869 protein sequences of the “PETase core domain” linked by 318,773 edges. The protein sequences depicted here were selected by clustering at a threshold of 90% sequence identity. Edges (links) were selected at a threshold of 55% sequence similarity. Nodes are coloured according to their annotated source organisms, with Actinobacteria in red, Proteobacteria in blue, Fungi in cyan, Bacteroidetes in orange, other bacteria from the FCB group in yellow, Planctomycetes in green, and unknown bacteria coloured in white*. See Methods section for more details on the network layout. Supplementary figures are available in **Figures S2** and **S3**.

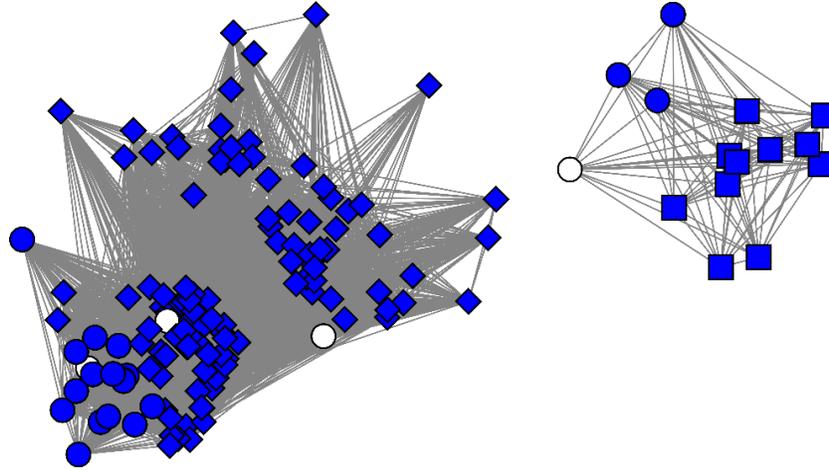


FIGURE 2. Network representation for 142 complete protein sequences similar to PURases linked by 6419 edges. The protein sequences depicted here were selected by clustering at a threshold of 90% sequence identity. Edges (links) were selected at a threshold of 60% global sequence similarity, without defining a core domain region. Nodes are coloured according to their annotated source organisms, with Proteobacteria in blue and unknown bacteria in white. The network on the left represents sequences with an N-terminal lid and a C-terminal β -sandwich domain and contains 127 nodes connected by 6314 edges. Diamonds represent sequences originating from the genus *Pseudomonas* (from the class Gammaproteobacteria). The network on the right represents sequences similar to carboxylesterases and contains 15 nodes connected by 105 edges. Squares represent sequences originating from the class of Betaproteobacteria. See Methods section for more details on the network layout.