PREMIO NACIONAL JUVENIL DEL AGUA 2022

SIWI Stockholm Junior Water Prize

Entry to Stockholm Junior Water Prize 2022

Discovery of Thermostable PET Degrading Enzymes from Metagenomes

Alonso Hernández Velázquez

MEXICO

TABLE OF CONTENTS

Discovery of Thermostable PET Degrading Enzymes from Metagenomes

Summary

By 2050, if immediate action is not taken, 12 billion tons of plastic will end up in landfills or in the environment, especially in water bodies, impacting the health and well-being of people and ecosystems.

Nowadays, traditional PET recycling and degradation technologies are slow and polluting. An alternative for these technologies is the use of enzymes produced by microorganisms, however, their sensitivity to high temperatures limits their application in industry and bioremediation.

By analyzing about a million protein sequences found in metagenomes from hydrothermal sites, this project discovered two thermostable PET-degrading enzymes that can be used in industry.

Key Words

PET (Polyethylene Terephthalate), PETase, METase, metagenomes, Hidden Markov Models, thermo-stability

Abbreviations and acronyms

PET	Polyethylene Terephthalate	UNAM	Universidad Nacional Autónoma de México
MHET	Mono-(2-hydroxyethyl)terephthalic acid	pH	Hydrogen Potential
IMG/MER	Integrated Microbial Genomes & Microbiomes	BLAST	Basic Local Alignment Search Tool
HMM	Hidden Markov Model	MUSCLE	Multiple Sequence Comparison by Log Expectation
USCD	University of California San Diego	SDGs	Sustainable Development Goals
NCBI	National Center for Biotechnology Information	CD-hit	Cluster Database at High Identity with Tolerance
ADN	Deoxyribonucleic acid	MEGA	Molecular Evolutionary Genetics Analysis
ICML	Institute of Marine Sciences and Limnology	PFAM	Protein Families database
BLASTp	Basic Local Alignment Search Tool Proteins	PAE	Predicted Aligned Error
RESB PDB	Research Collaboratory for Structural Bioinformatics Protein Data Bank	PCR	Polymerase Chain Reaction
MMseqs2	Many-against – Many sequence searching		

Acknowledgments

I would like to thank my first advisor MSc Rodrigo Hernández Velázquez, and my second advisor PhD Justo Salvador Hernández Avilés. I also appreciate the services provided by the "Chihuil" server that is part of the scientific infrastructure acquired through the project " Consolidación de las capacidades bioinformáticas en la Unidad Mazatlán del ICMyL-UNAM " No. 300659, financed by Fondo de Infraestructura 2019 of the Consejo Nacional de Ciencia y Tecnología CONACYT.

Short Bio

My name is Alonso Hernández Velázquez, graduated from Escuela Nacional Preparatoria #2, resident of Mexico City. Some of my hobbies are playing sports, enjoying a soccer game and studying philosophy. I have a great passion for scientific knowledge and water technologies. I consider myself a friendly person who enjoys diverse people.

INTRODUCTION

Currently, the pollution of water, particularly of the ocean, is one of the biggest problems that humanity is facing. One of the pollutants with a higher impact in the destruction of the ocean is plastic. It is estimated that every year, around 300 million tons of plastic are produced, and about 14 million tons end up in the ocean (The International Union for Conservation of Nature, 2021). According to a report by the World Economic Forum, if the production of plastics continues as of now, by 2050 there will be more plastic in the ocean than fish (Schwab, 2019). This, in addition to representing an environmental crisis, also has implications for human health due to the toxicity of plastics. It is now known that microplastics <10 mm can be transported to the circulatory system and accumulate in organs such as the liver, kidney, and brain (Vethaak & Legler, 2021). Furthermore, these microplastics are possible vectors of pathogenic microorganisms (Vethaak & Legler, 2021).

Polyethylene Terephthalate (PET), Environmental Impact and Degradation

One of the most used plastics worldwide is PET, and it is estimated that around 85 million metric tons are produced per year, which are mainly used to produce bottles, of which only between 15% - 35% are recycled (The National Renewable Energy Laboratory, 2021.)

Nowadays, in addition to traditional mechanical and chemical technologies that are usually slow and more polluting, new proposals have emerged to recycle and degrade PET using enzymes produced by microorganisms. Such is the case of *Ideonella sakaiensis*, a bacterium isolated from the sediment that has a system of two

enzymes: PETase and MHETase. These two enzymes catalyze the reactions for the degradation of PET (polyethylene terephthalate) to MHET (mono(2-hydroxyethyl) terephthalic acid) and subsequently, to terephthalic acid and ethylene glycol (Yoshida et al., 2016). Of note, these last two monomers are considered environmentally benign (Yoshida et al., 2016.).

One of the main limitations of these enzymes is that they carry out the reaction at temperatures of ~30°C. This problem is shared by most plastic-degrading enzymes, since they are generally highly sensitive to high temperatures (Purohit et al., 2020). One of the main characteristics of PET is its high ordered crystalline content, which reduces the mobility of the polymer chains and makes it less susceptible to enzymatic hydrolysis (Brott et al., 2021). This susceptibility increases considerably when the temperature is close to at the glass transition temperature of PET, which is higher than 70 °C (Alves et al., 2002.)

What is a Metagenome?

Since most microorganisms are not cultivable, one way to access the potential that exists in nature to degrade any contaminant is metagenomics. Metagenomics consists of extracting DNA from an environment for subsequent sequencing and gene prediction. From the genes it is possible to obtain the amino acid sequences of the proteins, which in turn allows functional annotations to be obtained using methods such as hidden Markov models (HMM) or alignments.

OBJECTIVES

- To discover and identify thermostable enzymes produced by microorganisms in natural environments, for their possible use and industrial application to help to reduce plastic pollution in water bodies.
- To build an HMM to identify PETases and METases in metagenomes and determine the denaturation temperature and stability of the enzymes found.

METHODS

Data Collection

Metagenomes were selected from a search in the IMG/MER webpage aided by a bibliographic review that took into account the following characteristics:

1. That they were available in the public IMG/MER database (Chen et al., 2021; Mukherjee et al., 2021), with a valid and publicly accessible identifier in accordance with the policy of the Joint Genome Institute, and that had been processed with the Metagenome Annotation Pipeline (MAP v.4) (Huntemann et al., 2016).

2. All sampling sites must have high temperatures ≥ 60°C (Figure 1).

3. Preferably, sampling sites with high anthropogenic activity, sites located on the seabed or ecosystems known for their high diversity.

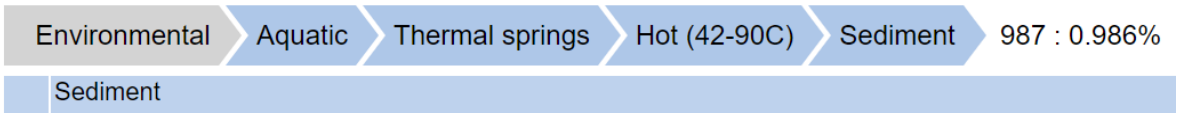

Figure 1. Example of metagenome search according to the characteristics of the sampling site.

To download and analyze the metagenomes, the server: "Chihuil the bioinformatics server" of ICML-UNAM was used. The script: jgi-query.py script was used to download the metagenomes from the IMG/MER database. Adding the genes of all the metagenomes (Table 1) almost a million (991,588) were obtained, it is worth mentioning that these genes also have the amino acid sequences for which they code.

TaxonOID	Country or Region	pH	Temperature °C	Latitud	Longitude	Number of Genes
3300014869	Mexico	-	~ 60	27°23′ 16.8″ N	111°33′ 36″ W	404,072
2007309000	USA	9.3	86	44°33′ 37.1″N	110° 50′ 1.804″ W	14,684
3300000859	China	9.3	82.5	24° 57′ 3.208″ N	98° 26′ 10.921″ E	16,654
3300000865	China	7.29	73.8	25°26′ 24.4″N	98°26′ 26.9″E	92,136
3300005095	Pacific Ocean	5.2	231	9°46′ 16.3″N	104°16′ 44″W	464,042
3300014954	Sweden	-	-	57° 43′ 12.251″ N	12° 3′ 2.042″ E	183,809
3300003879	Atlantic Ocean	-	22.4	35° 33′ 6.012″ N	65° 39′ 29.988″ W	489,522

Table 1. Metagenome data, including identifier (TaxonOID), sample location (region and coordinates), pH, temperature, and number of genes. Underlined in red negative control and in green positive control.

Controls

In order to have a data set where it is expected not to find sequences of PETases and MHETases, the genes of the microbiome from the fecal sample of a 12-month-old infant from Sweden were obtained. This control was used in a similar way to the study by Zrimec et al., 2021, taking into account that no plastic-degrading enzymes have been reported in the human microbiome. As a positive control, that is, a data set in which sequences of PETases and MHETases were expected to be present, a metagenome from plastic debris obtained from the surface of the Atlantic Ocean was used (Figure 2.)

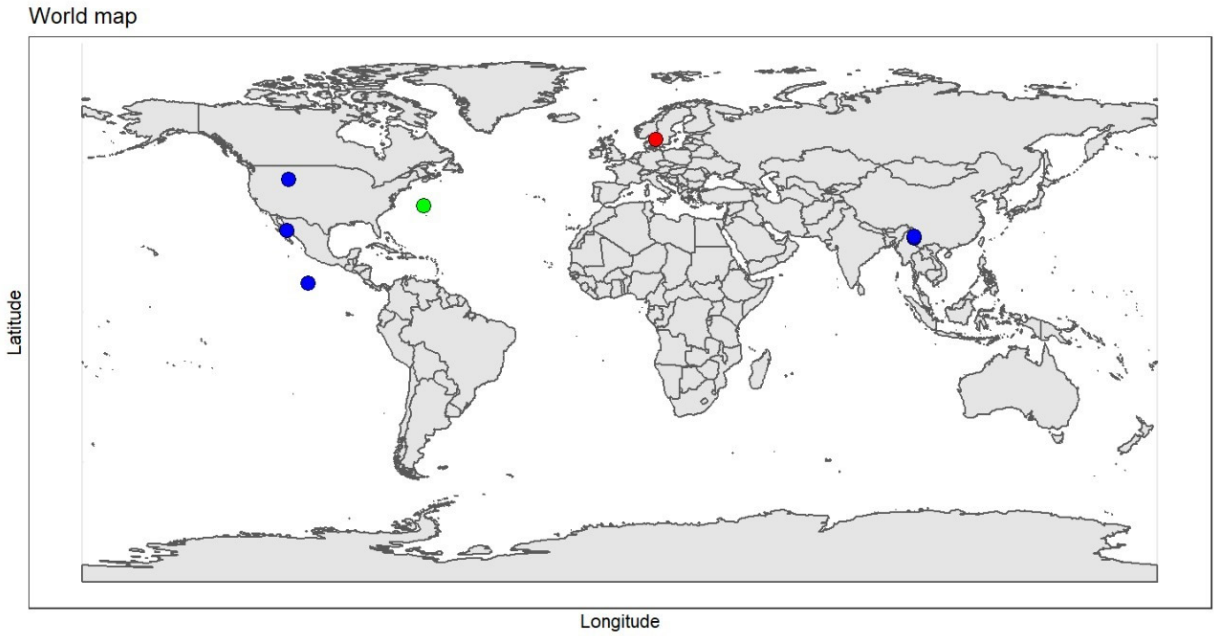

Figure 2. Map with coordinates of metagenome samples: hydrothermal metagenomes in blue, negative control in red, and positive control in green.

HMM Generation

Gene Download

Ideonella sakaiensis PETase and MHETase amino acid sequences were downloaded from NCBI: PETase and MHETase.

Search for Homologous Sequences with BLAST

Online BLAST (Johnson et al., 2008) was run, using PETase and MHETase sequences against the “Non-redundant protein sequences (nr)” database (Pruitt et al., 2005).

The results were filtered for PETase using a percentage identity between 60 and 100% and a coverage between 70 and 100%. In total, 48 homologous sequences were obtained. Results were similarly filtered for MHETase using percent identity between 50 and 100%, and coverage between 70 and 100%. In total, 17 homologous sequences were obtained.

Obtaining Representative Sequences with CD-HIT

CD-HIT (Fu et al., 2012) was used on the UCSD web server (Huang et al., 2010), with the set of PETase and MHETase sequences previously obtained with BLAST as an input. “Sequence identity cut-off” was the only modified parameter, where a percentage of 95% was used. In total, 22 representative sequences of PETases and 9 of MHETases were obtained.

Multiple Alignments

The multiple alignments from the selected sequences with CD-Hit were carried out with the MUSCLE program (Edgar, 2004) version: 3.8.1551. A maximum iteration value of 100 was used to ensure that the best possible alignment was obtained.

When reviewing the alignment of the PETases in MEGA v6 (Tamura et al., 2013), it was verified visually that the last two sequences caused gaps of ~138 amino acids. For this reason, these sequences from Burkholderia bacteria were eliminated.

Subsequently, the alignments were carried out with MUSCLE (Edgar, 2004) as previously described and the first 13 and the last 22 residues that practically consisted of their totality of gaps were removed.

Creating HMMs with the HMMER Pipeline

To create the Hidden Markov Models (HMMs) for each protein family, the alignments created with MUSCLE were input to hmmbuild the HMMER 3.3 pipeline (Finn et al., 2011) default parameters were used.

HMMS Validation

To validate the generated HMMs, the online version of hmmersearch (Finn et al., 2011) was used, which uses an HMM and searches it against a protein database. For this, the “Reference proteomes” database (Dessimoz et al., 2012) was selected, which contains manually and algorithmically curated high-quality proteomes, that span the tree of life. As cut-off e-value, 1x10-5 was used.

Hmmer vs Metagenomes

With the validated hidden Markov models, the search for PETases and MHETases in the metagenomes was carried out, for this hmmsearch (Finn et al., 2011) was used with the default parameters. From this search, the hits with e-value for the complete sequence ≤ 1e-5 were selected for each metagenome.

Characterization of Hits of PETases and MHETases

For each significant hit, the PFAM annotation (Punta et al., 2012) available from the Metagenome Annotation Pipeline (Huntemann et al., 2016) was searched. Subsequently, the amino acid sequences of each protein were used to perform a BLAST

search (Johnson et al., 2008) using the REFSEQ database (Pruitt et al., 2005) and not redundant database. For each search, the best hit that had an annotation was selected, from this hit the name of the organism to which it belonged was also obtained, in addition to its taxid.

Three-Dimensional Structure Modeling with Alphafold

The amino acid sequences of the proteins were used to obtain the three-dimensional structure using AlphaFold2_advanced, which is available online. Alphafold is a program that uses “machine learning” to obtain accurate three-dimensional predictions of proteins even when they do not have similar structures (Jumper et al., 2021). AlphaFold2_advanced was run in ColabFold (Mirdita et al., 2022) with the default parameters and the homooligomer option enabled, usingMMseqs2 (Steinegger & Söding, 2017) as the engine for the multiple alignment method.

Prediction of the Denaturation Tm of PETases and MHETases

To predict the temperature at which proteins denature, the Tm Predictor software (Ku et al., 2009) was used, which uses amino acid sequences as a reference, since certain amino acids confer greater stability, and a high proportion of these indicates higher denaturation temperature. Tm predictor provides two values: the Tm that corresponds to the temperature at which about 50% of the protein is unfolded and the Tm index (TI), which is a coefficient that if it is greater than 1 indicates a high stability. It is worth mentioning that Tm predictor reports the Tm in one of three categories: >65°C, 55°C~65°C, <55°C. Tm predictor was run on the online server (Ku et al., 2009) for PETase and MHETase from I. sakaiensis as a reference, in the same way it was run for putatively thermostable PETase and MHETase obtained from metagenomes

Prediction of MHETase Stability with FoldX

To ensure that the MHETase from bacterium HR28 obtained from a metagenome was more stable than that from I. sakaiensis, FoldX (Schymkowitz et al., 2005) for Windows was used, which allows, from the three-dimensional structure in pdb format, to predict the stability of a protein. Therefore, this program was run with the default parameters for the three-dimensional structure of the thermostable MHETase obtained with ColabFold (Mirdita et al., 2022) and the three-dimensional structure of I. sakaiensis obtained from the RCSB PDB database (Berman et al., 2000)

RESULTS

HMMs Validation

The result of the validation of the HMM of the MHETases was positive, since the highest hits were also all annotated as: “Tannase and feruloyl esterase” or “Feruloyl esterase”. Tannases are hydrolases that act on carboxylic esters. MHETases belong to this family of proteins and evidence indicates that they evolved from ferulic acid esterases (Knott et al., 2020).

When analyzing the domains of the results, it was found that they were extremely consistent since all of the 4205 sequences that were annotated with this profile, had a domain of the tannase family (Figure 3).

Figure 3: Domain of the Tanase family, present in the sequences annotated with the HMM of MHETase

For PETases, a total of 910 results were obtained. The annotations of the main results are reported in Table 2. All these domains correspond to the function of the PETases, since they are hydrolases that, according to the three-dimensional structures available, have the domain: α/β- hydrolase fold with the difference that the active site is more open than in cutinases (Knott et al., 2020).

Domain	Number of sequences
Platelet-activating factor acetylhydrolase, isoform II (PAF-AH_p_II)	267
Dienelactone hydrolase (DLH)	209
no domain architecture	162
Chlorophyllase2	137
Hydrolase_4	63

Table 2. Domains found when searching the HMM of PETases against the “Reference proteomes” database.

The representative sequence of proteins without a domain was searched with BLASTp (Johnson et al., 2008) against the Uniprot database (The Uniprot Consortium, 2021) and the closest sequence with a percentage of identity of 61.5% was the PETase of *Ideonella sakaiensis* (Figure 4). Which indicates that these sequences correspond to PETases even though they do not have an assigned hydrolase domain.

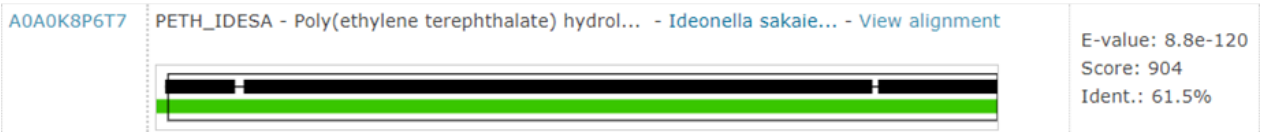

Figure 4. Search result of representative domain-free protein sequence with BLASTp against UniProt database. The e-value, Score and Percentage of identity (Ident) are reported.

PETases and MHETases in Metagenomes

The metagenome from “Hot springs sediment microbial communities from Tengchong” was the only one for which a PETase and a MHETase were found (Table 3), which suggests that in this system the degradation of PET to terephthalic acid and ethylene glycol is carried out. When searching for this PETase with BLASTp, a hit with 100 coverage and 98.5% identity corresponding to the organism Candidatus Kryptobacter tengchongensis was obtained, while for the MHETase a hit with 100 coverage and 96.09% identity corresponding to the bacterium HR28 organism was obtained. It is worth mentioning that both organisms are thermophiles.

This metagenome was obtained in China from a thermal pool that visibly has human activity and possibly contamination (Figure 5).

Figure 5 Gongxiaoshe hydrothermal pool in Tengchong, China (A), PET bottles and plastic debris found next to hot spring (B). Source: Joint Genome Institute, 2016

It should also be mentioned that no PETase or MHETase sequence was obtained in the negative control, while a total of 7 PETases were obtained in the positive control (Table 3).

ID	Sample	Country or Region	Number of PETases	Number of MHETases
2007309000	Yellowstone Bath Hot Springs	USA	0	0
3300000859	Hot spring sediment from Guminquan	China	0	0
3300000865	Hot springs sediment microbial communities from Tengchong	China	1	1
3300003879	Plastic marine debris microbial communities from the Atlantic Ocean	Atlantic Ocean	7	0
3300005095	Hydrothermal chimney microbial communities from East Pacific Rise	Pacific Ocean	0	2
3300014869	Subseafloor sediment microbial communities from Guaymas Basin,	Mexico	0	0
3300014954	Human fecal microbial communities from infant at 12 months	Sweden	0	0

Table 3: JGI IDs, description of the metagenome sample, country or location where the sample was collected, and number of PETases and MHETases. The positive control is underlined in green and the negative control in red.

Three-Dimensional Structures

From the three-dimensional prediction, it becomes clear that the structures have a more compact folding than those of I. sakaiensis, suggesting greater stability. The predictions also presented high values of pLDDT, which estimates the confidence of each residue (0-100) and allows to assess locally the quality of the structure. As shown in Figure 6 (A and B), for both proteins almost all of the aminoacids have a very high pLDDT value (>90%). The only exception corresponds to few amino acids at the edges of proteins which is common in Alphalfold and does not affect the overall structure. In the case of PETase, there is an alpha helix that corresponds to a transmembrane domain (analysis not shown), which has lower values in a couple of residues (Figure 6B). This is because Alphafold is not ideal for these types of structures, which however do not affect enzyme function. Another important metric to consider is the PAE (Predicted Aligned Error). This metric is used to predict the confidence between pairs of amino acids in the protein, that is, how certain we are that an amino acid is in the correct position considering the position of another one in the structure. PAE values can be represented with a heatmap (Figure 6. C, D) and in the case of PETase (Figure 6D) only the error values corresponding to the transmembrane domain are high, which is again due to the characteristics of Alphafold. In the case of the MHETase, almost all values are close to zero, indicating that all domains were correctly assigned with respect to each other (Figure 6C).

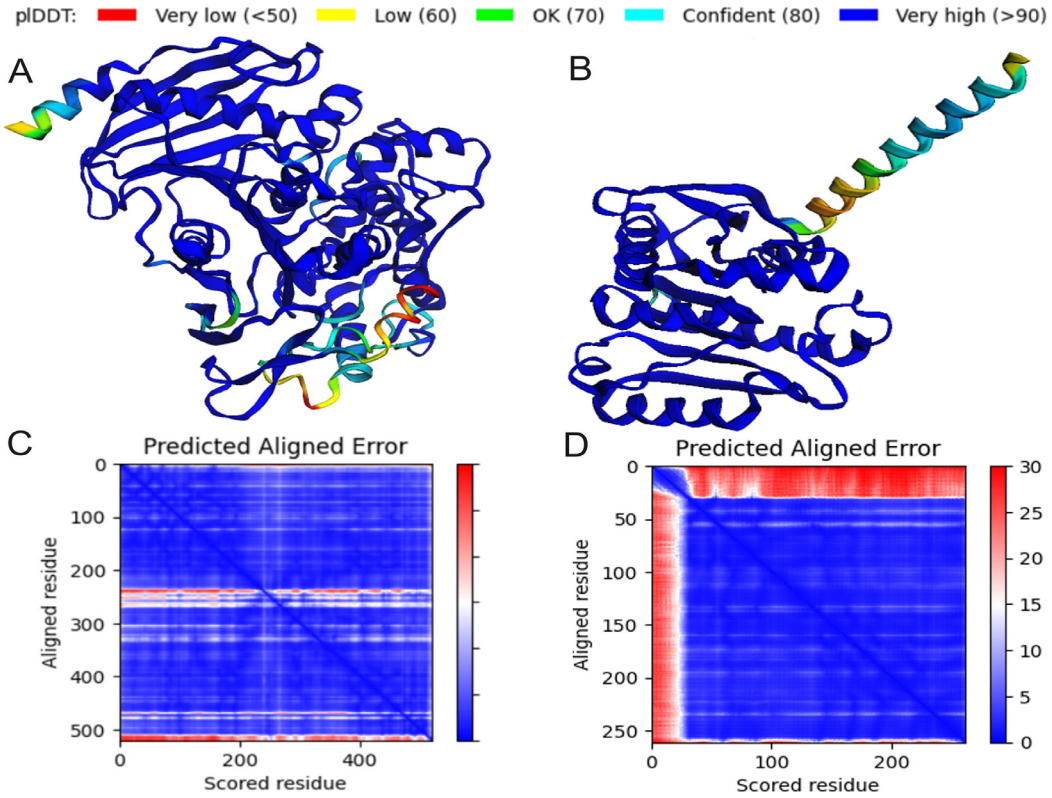

Figure 6. Three-dimensional structure and pDDT values for MHETase (A) PETAse (B). PAE values for MHETase (C) and PETase (D)

Denaturation Temperature of PETase and MHETase

The thermostable PETase of Candidatus K. Tengchongensis is clearly more stable than that of I. sakaiensis. According to Tm predictor its denaturation temperature (Tm) is >65°C, while that of I. sakaiensis is 55°C~65°C . On the other hand, although the Tm index (TI) of thermostable MHETase is close to 1 (0.92530989053), it falls into the same Tm category as I. sakaiensis MHETase (55°C~65°C) (Table 4). This may be because the stability of the bacterium HR28 thermostable MHETase stems from its three-dimensional structure, which is not captured directly from its amino acid structure.

Enzyme	Denaturation Temperature (Tm)	Tm Index(TI)
PETase *I. sakaiensis*	55°C~65°C	0.771523898661
PETase *Candidatus K. tengchongensis*	>65°C	1.18698424548
MHETase *I. sakaiensis*	55°C~65°C	0.767726274251
MHETase bacterium HR28	55°C~65°C	0.92530989053

Table 4. Denaturation temperatures according to the three predictor Tm categories and Tm indices. Thermostable enzymes obtained from metagenomes are indicated in red.

Stability of the MHETase from Its Three-Dimensional Structure

FoldX stability results are expressed as the Gibbs free energy (ΔG) of unfolding (kcal mol−1) and take into account various factors. In the case of the MHETase from I. sakaiensis, its free energy is 92.76 kcal mol−1, while that of the thermostable MHETase from bacterium HR28 is 315.84 kcal mol−1. This means that to denature this protein more than triple the energy is required, being therefore more stable.

DISCUSSION

The Hidden Markov models were generated and validated and showed to be consistent in the identification of PETases and MHETases. This represents a valuable resource that can be used in even larger data sets for the identification of this type of enzymes in the metagenome of any environment or in the genome of any organism.

Two enzymes capable of degrading PET plastic (polyethylene terephthalate) to MHET (mono(2-hydroxyethyl) terephthalic acid) and later to terephthalic acid and ethylene glycol were discovered. From the PETase and MHETase that were found in the thermal pool of Tengchong China, the three-dimensional structure was obtained, and it was determined that their thermal stability was superior to that of *Ideonella sakaiensis*. Using bioinformatic methods, it was found that for the PETase, the Tm is greater than 65 °C and that the MHETase of the metagenome has a ΔG three times greater than that of I. sakaiensis. Since the temperature of the sampling site in Tengchong was 73.8 °C, it is assumed that these enzymes withstand equivalent temperatures, since these types of enzymes are generally exported to the outside of the cell.

One of the limitations of the I. sakaiensis PETase is that it loses its activity at 40 ºC after 24 hours (Son et al., 2019). The study presented by Son et al., 2019 shows with thermostable PETases obtained using protein engineering, that thermostability is the crucial characteristic for the efficiency of PET degradation, in this study, the thermostable PETases used carried out the reaction for 120 hours at 40 ºC and were denatured after 1 hour at 50 ºC. These results suggest that the thermostable enzymes found in the present study from the Tengchong metagenome at 73.8 °C, will have a better performance, and will carry out the reaction for a longer time.

This last characteristic makes these enzymes ideal for industrial use, where PET is washed and treated at temperatures around 75 °C. This temperature also corresponds to the glass transition temperature of PET (Alves et al., 2002), which makes it more susceptible to being hydrolyzed.

This system of enzymes can be used with the aim of closing the plastic cycle, which involves the depolymerization of PET, to later generate value-added products with the monomers and thus develop a circular economy (Kakadellis & Rosetto, 2021).

Similarly, using these enzymes in objects made from PET, such as bottles, would increase their malleability, which would reduce recycling costs. It is worth mentioning that recycling plastics consumes 88% less energy than producing them from their raw materials (Johnson & Lichtveld, 2022), in this way using enzymes that reduce processes such as PET grinding would bring PET recycling closer to carbon neutrality.

Regarding bioremediation, since terephthalic acid and ethylene glycol are highly biodegradable (Yoshida et al., 2016), this two-enzyme system is essential to recover natural environments, taking into account that ~40% of the plastic is found in these ecosystems (Worm et al., 2017). An area of main importance is the garbage island known as the "Great Pacific garbage patch", which covers an area of 1.6 million square kilometers and has around 1.8 trillion pieces of plastic (Lebreton et al., 2018). In this site, the thermostable PETase and MHETase from Tengchong could be used due to their expected longer half-life given their high denaturation temperature.

An important aspect for the use of thermostable PETase and MHETase is the expression of these enzymes in recombinant systems. Said expression can be carried out in model organisms such as Escherichia coli for subsequent purification, as is done with other enzymes such as taq polymerase, which has optimal activity at temperatures of 75-80°C, and is used in PCR. (Moazen et al., 2012). Another alternative is to express these enzymes in thermophilic Bacillaceae (Drejer et al., 2018), these microorganisms could be used directly on the plastic at high temperatures.

Considering that 51% of world plastic production is concentrated in a limited number of industrialized countries (PlasticsEurope, 2021), and that these countries are making investments in innovative recycling technologies, it is a reality that the use of enzymes and microorganisms will be more and more common. In this context, thermostable PETases and MHETases will play a fundamental role, considering that the PET produced is equivalent to 12% of the total volume of solid waste worldwide (George & Kurian, 2014).

CONCLUSIONS

Two thermostable enzymes with the ability to degrade PET plastic to its environmentally friendly monomers were discovered. One of these enzymes is a PETase and the other a MHETase, both found in the metagenome of a hot spring in Tengchong China. The superior thermostability of PETase and MHETase from Tengchong with respect to those from *Ideonella sakaiensis* was also determined by bioinformatic approaches.

Other valuable resources were the hidden Markov models, which were validated and shown to be consistent in the identification of PETases and MHETases and can

also be used to search for these enzymes in other genomes and metagenomes.

This study to date represents the only effort to obtain sequences of thermostable PET degrading enzymes from thermal zones, the most used alternatives consist of using mutagenesis techniques or directed evolution to increase the stability of the enzyme. These strategies, in addition to being very expensive due to the need to use molecular biology techniques, are time-consuming and waste the potential of microorganisms to degrade practically any substrate on earth. Likewise, this study serves as a precedent to employ similar approaches for other types of plastics and pollutants, considering that there are thousands of metagenomes that can be explored for the discovery of new enzymes.

Finally, to use these enzymes on an industrial scale, pilot tests must be carried out to experimentally determine their catalytic activity under different conditions, which is why alliances with biotechnology companies and plastic recycling plants are necessary.

Contributions

This project seeks to influence the achievement of Sustainable Development Goal 6 “Clean water and sanitation”; in particular, of targets 6.3 and 6.6. As well as the achievement of SDG 3 “Health and well-being”, in its goal 3.9, which specifies the urgency of reducing mortality and diseases associated with the presence of dangerous chemical products in water and soil, by contributing to the development of new technologies. to improve water quality by reducing pollution caused by PET and microplastics, minimizing the presence of hazardous products and materials.

It would also make it easier for different industries to manage waste responsibly and thereby advance SDG 12 “Responsible production and consumption”, in particular target 12.4 by favoring the rational management of PET and reducing the release of harmful substances in water bodies.

Bibliography

Alves, N. M., Mano, J. F., Balaguer, E., Meseguer Dueñas, J. M., & Gómez Ribelles, J. L. (2002). Glass transition and structural relaxation in semi-crystalline poly(ethylene terephthalate): a DSC study. Polymer, 43(15), 4111–4122.Brott, S., Pfaff, L., Schuricht, J., Schwarz, J.-N., Böttcher, D.,

Badenhorst, C. P. S., Wei, R., & Bornscheuer, U. T. (2021). Engineering and evaluation of thermostable IsPETase variants for PET degradation. *Engineering in Life Sciences*, *22*(3–4), 192–203.

Chen, I.-M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., Hajek, P., Ritter, S., Varghese, N., Seshadri, R., Roux, S., Woyke, T., Eloe-Fadrosh, E. A., Ivanova, N. N., & Kyrpides, N. C. (2021). The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. Nucleic Acids Research, 49(D1), D751–D763.

Dessimoz, C., Gabaldón, T., Roos, D. S., Sonnhammer, E. L. L., Herrero, J., & Consortium, the Q. for O. (2012). Toward community standards in the quest for orthologs. Bioinformatics, 28(6), 900–904.

Drejer, E. B., Hakvåg, S., Irla, M., & Brautaset, T. (2018). Genetic Tools and Techniques for Recombinant Expression in Thermophilic Bacillaceae. Microorganisms, 6(2).

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, *39*, 29–37.

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics, 28(23), 3150–3152.

George, N., & Kurian, T. (2014). Recent Developments in the Chemical Recycling of Postconsumer Poly(ethylene terephthalate) Waste. Industrial & Engineering Chemistry Research, 53(37), 14185–14198.

Huang, Y., Niu, B., Gao, Y., Fu, L., & Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, *26*(5), 680–682.

Huntemann, M., Ivanova, N. N., Mavromatis, K., Tripp, H. J., Paez-Espino, D., Tennessen, K., Palaniappan, K., Szeto, E., Pillay, M., Chen, I.-M. A., Pati, A., Nielsen, T., Markowitz, V. M., &

Johnson, B. L., & Lichtveld, M. Y. (2022). Environmental Policy and Public Health: Emerging Health Hazards and Mitigation, Volume 2. Taylor & Francis Limited.

Joint Genome Institute. (2016). Gongxiaoshe Spring in China with rural Chinese village of Dientan in background by Hedlund [Photograph]. https://jgi.doe.gov/wp-content/uploads/2016/01/1K-Gongxiaoshe-Spring-in-China-with-rural-Chinese-village-of-Dientan-in-background-by-Hedlund.jpg

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, *36*, 5-9.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589.

Kakadellis, S., & Rosetto, G. (2021). Achieving a circular bioeconomy for plastics. Science, 373(6550), 49–50.

Kyrpides, N. C. (2016). The standard operating procedure of the DOE-JGI Metagenome Annotation Pipeline (MAP v.4). *Standards in Genomic Sciences*, *11*(1), 17.

Knott, B. C., Erickson, E., Allen, M. D., Gado, J. E., Graham, R., Kearns, F. L., Pardo, I., Topuzlu, E., Anderson, J. J., Austin, H. P., Dominick, G., Johnson, C. W., Rorrer, N. A., Szostkiewicz, C. J., Copié, V., Payne, C. M., Woodcock, H. L., Donohoe, B. S., Beckham, G. T., & McGeehan, J. E. (2020). Characterization and engineering of a two-enzyme system for plastics depolymerization. *Proceedings of the National Academy of Sciences*, *117*(41), 25476–25485.

Ku, T., Lu, P., Chan, C., Wang, T., Lai, S., Lyu, P., & Hsiao, N. (2009). Predicting melting temperature directly from protein sequences. *Computational Biology and Chemistry*, *33*(6), 445–450.

Lebreton, L., Slat, B., Ferrari, F., Sainte-Rose, B., Aitken, J., Marthouse, R., Hajbane, S., Cunsolo, S., Schwarz, A., Levivier, A., Noble, K., Debeljak, P., Maral, H., Schoeneich-Argent, R., Brambini, R., & Reisser, J. (2018). Evidence that the Great Pacific Garbage Patch is rapidly accumulating plastic. Scientific Reports, 8(1), 4666.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nature Methods.

Moazen, F., Rastegari, A., Hoseini, S. M., Panjehpour, M., Miroliaei, M., & Sadeghi, H. M. M. (2012). Optimization of Taq DNA polymerase enzyme expression in Escherichia coli. Advanced Biomedical Research, 1, 82.

Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Sundaramurthi, J. C., Lee, J., Kandimalla, M., Chen, I.-M. A., Kyrpides, N. C., & Reddy, T. B. K. (2021). Genomes OnLine Database (GOLD) v.8: overview and updates. Nucleic Acids Research, 49(D1), D723–D733.

PlasticsEurope, 2021. Plastics-the Facts 2021. Available from: https://plasticseurope.org/knowledge-hub/plastics-the-facts-2021/

Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research, 33, D501–D504.

Purohit, J., Chattopadhyay, A., & Teli, B. (2020). Metagenomic Exploration of Plastic Degrading Microbes for Biotechnological Application. *Current Genomics*, *21*(4), 253–270.

Schwab, K. (2019). *The Global Competitiveness Report 2019: World Economic Forum.* Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., & Serrano, L. (2005). The FoldX web server: an online force field. *Nucleic Acids Research*, *33*, 382–388.

Son, H. F., Cho, I. J., Joo, S., Seo, H., Sagong, H.-Y., Choi, S. Y., Lee, S. Y., & Kim, K.-J. (2019). Rational Protein Engineering of Thermo-Stable PETase from Ideonella sakaiensis for Highly Efficient PET Degradation. ACS Catalysis, 9(4), 3519–3526.

Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nature Biotechnology, 35(11), 1026–1028.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725–2729.

The International Union for Conservation of Nature (IUCN). (2021, November). Marine plastic pollution. https://www.iucn.org/resources/issues-briefs/marine-plastic-pollution

The National Renewable Energy Laboratory. (2021, October). Researchers Engineer Microorganisms To Tackle PET Plastic Pollution. https://www.nrel.gov/news/program/2021/researchers-engineer-microorganisms-to-tackle-pet-plastic-pollution.html

The UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*, 480–489.

Vethaak, A. D., & Legler, J. (2021). Microplastics and human health. *Science*, *371*(6530), 672–674.

Yoshida, S., Hraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., Toyohara, K., Miyamoto, K., Kimura, Y., & Oda, K. (2016). A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science*, *351*(6278), 1196–1199.

Zrimec, J., Kokina, M., Jonasson, S., Zorrilla, F., & Zelezniak, A. (2021). Plastic- Degrading Potential across the Global Microbiome Correlates with Recent Pollution Trends. *MBio*, *12*(5), 2155-21.