BIOCHEMICAL AND BIOPHYSICAL CHARACTERIZATION OF PLASTIC DEGRADING AROMATIC POLYESTERASES

by

Ece Topuzlu

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Biochemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

November 2019

©COPYRIGHT

by

Ece Topuzlu

2019

All Rights Reserved

DEDICATION

To my mother and my father, to whom I owe *everything*.

ACKNOWLEDGEMENTS

I wouldn't be where I am today if it weren't for the love, help and support of my mentors, advisors, collaborators, friends and family. I would first like to express my deepest, unending gratitude for my advisor, Valérie Copié, for her deeply empathetic nature, for believing in me, for her affable, but fierce personality, and her support throughout the years. I am thankful for the insightful discussions and comments my graduate committee members Prof. Brian Bothner, Prof. Martin Teintze and Prof. Jen DuBois have provided. Finally, I'd like to thank Dr. Gregg Beckham for opening his doors for me at the National Renewable Energy Laboratory. Conception of this thesis at its current form would not have been possible without him. The professional, supportive and excellent research environment he let me be a part of has broadened my horizons to an extent I didn't think was possible. I'm also greatly indebted to Dr. Bryon Donohoe for his time and effort on my project. I would especially be lost without the help of Doreen Brown.

I'm incredibly lucky for having my brother Onur, my sister in law, Arzu, my niece and nephew Ada and Aras being here in the United States. I'm thankful for my amazing boyfriend, Justin, and his seemingly unending political rants for reminding me why I need to contribute more to this world. Finally, and most importantly, I'd like to thank my mother Nilgün, and my father, Celal for being the bravest people I know, and letting both of their children go to faraway lands in search for a better future. Here's to hoping that one day achieve the Sisyphean task of becoming a tiny fraction of the person they think I am today.

TABLE OF CONTENTS

0. SCOPE OF THIS WORK 1

1. SYNTHETIC POLYMER POLYETYHELENE TEREPHTHALATE AND ITS DEGRADATION MECHANISMS 2

Introduction to Synthetic Polymers and Polyethylene Terephthalate 2
Synthetic Polymers as Environmental and Health Hazards 4
Recycling of PET 6
Biodegradable and Biobased Plastics 7
Enzymatic depolymerization of PET 9
A two-enzyme system for PET hydrolysis in *Ideonella sakaiensis* 12
PETase is an α/β Hydrolase 13
Surface Degradation Capabilities of PETase 18
Structural and Biochemical features of MHETase 20
Significance of This Work 21
Outline of Work with Polyesterases 23

2. MATERIALS AND METHODS 24

Introduction 24
Solid Substrate Assays 24
Addition of Surfactants to Solid Substrate Assays 25
Crystallinity Measurements of Polymer Samples 25
Scanning Electron Microscopy 26
Ultramicrotomy of PET and PEF 27
Expression and Purification of Proteins 28
Small Scale Screening for Soluble Protein Expression 29
Specific Approaches to Protein and Expression and Purification 32
01-PETlid, 04-MHETCC2WS 32
06-PETCC 34
02-MHETnoLid, 12-MHETnolidCCwtPET, 13-MHETnolidCCdmPET 35
05-MHETCC2HF, 07-MHETCat, 14-MHET131Gly, 15MHET495Ile, 17-MHETlip, 16-MHET6CCFae, 19-MHET7CC 36
Putative MHETase from *Comomonas thiooxidans* 37
Putative MHETase from *Hydrogenaphaga sp.* 38
Kinetic Assays 39
p-NO Acetate, *p*-NP Butyrate and *p*-NP Octanoate Assays 40
Testing for Interfacial Activation of MHETase 40
MHETase & MHET Assays 40

TABLE OF CONTENTS CONTINUED

MHETase & BHET Assays 41
MHETase & Methyl Ferulate Assays 41
PETase & BHET Assays 41
PETase & MHETase Synergy 42

3. BIOCHEMICAL AND SURFACE CHARACTERIZATION OF HYDROLYSIS BY PETASE 45

Visualizing PETase Activity 45
Preferential Hydrolysis by PETase 46
Hydrolysis of Various Substrates by PETase 50
Unraveling Nano-scale Features of PET Deconstruction Mechanisms by PETase 53
Effect of Temperature and Various Additives on PET Degradation 57
Temperature 57
Surfactants 59
PETase Activity on Non-Native Substrates 60
Hydrolysis of *p*-NP Esters 60
Hydrolysis of BHET 62
MHETase activity on Non-Native Substrates 63
Exploratory Investigation of Putative PETases 65
Concluding Remarks 65

4. STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF MHETASE 68

Contribution of Authors and Co-Authors 68
Manuscript Information Page 70
Abstract 71
Significance 72
Introduction 72
Results 76
Structural Characterization of MHETase 76
Bioinformatics Analysis of MHETase Sequences and Structural Homology 79
Biochemical Characterization and Engineering 81
Molecular Simulations and Docking 81
MHETase synergistic interaction with PETase 86
MHETase activity on Non-Native Substrates 87
Discussion 87
Conclusions 89

TABLE OF CONTENTS CONTINUED

Author Contributions 89
Acknowledgements 90
Supplementary Materials and Methods 91
Plasmid construction 91
Ligand Synthesis 104
MHET and iMHET Synthesis 104
Crystallography 105
Expression and purification of PET mutant MHETase mutants, and putative MHETases 108
Analysis of Michaelis-Menten Kinetics of MHETase Variants 109
Enzyme Incubation with Substrate 109
Scanning Electron Microscopy 110
PETase and MHETase Synergy 110
MHETase activity on MHEI 111
HPLC Method 112
Bioinformatics 115
Sequence Selection and Conservation Analysis 115
Phylogenetic Analysis 117
Molecular Docking 118
MHETase Molecular Dynamic Simulations 118
Ligand Structure Preparation 119
Flexible Ligand/Flexible Receptor Docking 120
Molecular Simulations 121

5. DISCUSSION AND CONCLUDING REMARKS 128

Concluding Remarks 135

APPENDICES 137
APPENDIX A: Development of 1H NMR Metabolomics Methods For Human Serum 138
APPENDIX B: Crenarchaeal Adaptive Immune System Function and Regulation by The Protein Csa3 144

REFERENCES CITED 164

LIST OF TABLES

Table Page

1-1 Biobased polymers 9

1-2 PETase structures deposited on PDB to date....14

2-1 Substrates used in this study....26

2-2 Proteins expressed and purified in this study 30

2-3 Proteins expressed and purified in this study continued....31

2-4 Results of the small-scale expression trials 32

2-5 Fit summaries of response surface equations....43

2-6 Equations for the quadratic model of response surface analysis of PETase and MHETase synergy....43

2-7 ANOVA of the quadratic equation terms of the response surface of PETase and MHETase synergy....44

3-1 Estimated lamellar thickness (*lc*) of lcPET 56

3-2 Estimated lamellar thickness of (*lc*) boPET 57

3-3 Kinetic parameters of hydrolysis of *p*-NP acetate and *p*-NP butyrate by WT PETase, W159H/S238F and W159A 61

3-4 Kinetic parameters of hydrolysis of BHET by WT PETase and W159H/S238F....64

4-1 MHETase mutant plasmid construction 92

4-2 Synthesized MHETase mutant DNA fragments....97

4-3 PCR primers used in this study 102

4-4 Data collection and structure refinement of MHETase 107

4-5 Fit summaries of equations of various orders tested for response surface 112

LIST OF TABLES CONTINUED

Table Page

4-5 Fit summaries of equations of various orders tested for response surface 112

4-6 Fit summaries for the quadratic model describing synergy....113

4-7 ANOVA for the quadratic model describing synergy 115

4-8 MHETase incubation with MHEI....116

A-1 Comprehensive list of metabolites identified in Chenomx....143

B-1 Sulfolobus strains mentioned and used in this project....157

B-2 Primers used in the spacer acquisition assays....162

LIST OF FIGURES

Figure Page

1-1 Structures of polyethylene terephthalate and its precursors 3

1-2 Plastics and their physical damage on marine fauna 4

1-3 Notable biobased plastics 10

1-4 Fold diagram of a canonical α/β hydrolase fold 11

1-5 PET hydrolysis by PETase and MHETase 12

1-6 Structure comparison of *I. sakaiensis* PETase and *T. fusca* cutinase 15

1-7 Active site clefts of *I. sakaiensis* PETase and *T. fusca* Cutinase compared 16

1-8 Catalytic triads and adjacent amino acids of *I. sakaiensis* PETase and *T. fusca* cutinase 17

2-1 Workflow for analyzing ultrathin sections of lcPET and boPET 28

3-1 Changes visible to the naked eye after treatment by PETase at 30°C for 96 hours 46

3-2 Surface pitting observed in lcPET by treatment at 30°C for 96 hours 47

3-3 Mass loss of lcPET coupons upon incubation with enzyme 48

3-4 Patterns of preferential hydrolysis observed on lcPET surface by PETase after incubation at 30°C· 49

3-5 Hydrolysis of sPET surface by WT PETase and W159H/S238F 50

3-6 Incubation of boPET with WT PETase and W159H/S238F 51

3-7 Hydrolysis of PET fibers by PETase 53

3-8 Representative TEM micrographs of RuO_4 stained boPET, pre and post-digestion by WT PETase 54

LIST OF FIGURES CONTINUED

Figure Page

3-9 TEM micrographs of lcPET and boPET, pre and post-digestion. 55

3-10 Effect of temperature on hydrolysis of PET by WT PETase, W159H/S238F and W159A 58

3-11 Effect of surfactants on lcPET mass loss....60

3-12 Hydolysis of short chain *p*-NP esters by WT PETase, W159A, and W159H/S238F....63

3-13 Initial hydrolysis experiment of BHET after 18 hours 64

4-1 Structural analysis of MHETase showing overall enzyme structure catalytic triad and calcium binding site 77

4-2 Active site configuration of molecular dynamics configurations and hydrogen bonding distances of active site residues with MHET 83

4-3 Catalytic mechanism of MHETase....84

4-4 Response surface of the synergistic action of PETase and MHETase on lcPET....86

4-5 ^{1}H NMR of MHET with peak assignments 105

4-6 Modeled 3D response surface of PETase and MHETase synergy 111

4-7 Homology of 6,671 Tannase family sequences compared to *Is*MHETase....118

4-8 Phylogenetic analysis of 120 Tannase Family sequences 119

4-9 Conservation analysis of 6,671 Tannase family sequences....121

4-10 Amino acid frequencies of important residues in MHETase 122

4-11 Disulfide bond cysteines in 6,671 Tannase family sequences....123

4-12 Bioinformatic analyses of calcium binding site residues in *Is*MHETase....125

LIST OF FIGURES CONTINUED

Figure Page

5-1 Structures of PETase and MHETase to scale 135

5-2 Structural alignment of PETase and MHETase.....136

A-1 Representative NMR spectrum from a serum sample 140

A-2 2D PCA plot of the metabotypes from t=0.....141

A-3 3D PCA plots of the metabotypes from t=0 141

A-4 2D PCA plot of the metabotypes from t=12 w.....142

A-5 3D plots of the metabotypes from t=12 w 142

B-1 Ribbon diagram of the *si*Csa3 dimer 151

B-2 Topology or fold diagrams of Csx3 fold, Ferredoxin fold and dinucleotide binding domain with Two Rossmann Folds.....155

B-3 Acquisition cassette of *S. islandicus* M 16.4 156

B-4 Spacer acquisition assay 157

B-5 Purification of Csa3 from *E. coli* PlysS cells.....160

B-6 Multiple, twinned Csa3 crystals emerging from a single nucleation point as well as single crystal Csa3 161

B-7 Published work from RNA Biology 163

GLOSSARY

BHET – bis (2-hydroxyethyl) terephthalate
EG – Ethylene glycol
DMSO – Dimethyl sulfoxide
IPTG – Isopropyl- β-D-1-thiogalactopyranoside
MD – Molecular dynamics
MHET – mono (2-hydoxyethyl) terephthalate
MHETase – MHET hydrolase from *Ideonella sakaiensis.*
PBS – Polybutylene succinate
PDB – Protein Data Bank
PEF – Polyethylene 2,5-furandicarboxylate
PET – Polyethylene terephthalate
PETase – PET hydrolase from *Ideonella sakaiensis*
PLA – Polylactic acid
SEM – Scanning electron microscopy
TEM – Transmission electron microscopy
*Tf*Cut – *Thermobifida fusca* cutinase
T_g. – Glass transition temperature
TPA – Terephthalic acid
WT – Wild Type
W159H/S238F – Double mutant PETase from *Ideonella sakaiensis*

ABSTRACT

As the world is producing more plastics than it can recycle, accumulation of man-made polymers in the environment is becoming one of the greatest global threats humanity is facing today. One of the major contributors to the plastics pollution problem is polyethylene terephthalate (PET), an aromatic polyester widely used in the packaging, beverage, garment and carpeting industries.

As a response to the onslaught of plastics in the environment, fungi and bacteria are evolving metabolic pathways to convert plastics into useable energy sources. One of these organisms, a bacterium, *Ideonella sakaiensis* 201-F6, has recently been identified to convert PET into its monomers, terephthalic acid (TPA) and ethylene glycol (EG), and to use these compounds for energy and growth. *I. sakaiensis'* ability to convert PET is made possible by two enzymes, named PETase and MHETase. As a first step, PETase breaks down the insoluble substrate PET into a soluble major hydrolysis product – mono-(2-hydroxyethyl) terephthalate (MHET), which is then further hydrolyzed by MHETase into TPA and EG. Crystal structure of PETase, as well as some of its biochemical features, have been reported several times to date, but MHETase has remained largely uncharacterized.

This work focuses on further discovery-driven biophysical and biochemical characterization of PETase, visualization of PETase activity on various polyester surfaces, as well as the structural and biochemical characterizations of the MHETase enzyme. We have found that several aspects of PETase-mediated substrate surface modification hydrolysis mechanisms differ depending on the specific mechanical and material characteristics of the substrate. We have also found that PETase is inhibited by BHET.

Additionally, we have solved the crystal structure of MHETase. MHETase consists of an α/β hydrolase domain, and a "lid" domain, commonly seen in lipases. Molecular dynamics simulations revealed the mechanism of MHETase action. Through bioinformatics approaches, we have also identified mutants of interest for improved MHETase activity. Coincubation of MHETase with PETase affects PET turnover in a synergistic fashion.

Taken together, this work provides additional insights into the mechanisms of action of the PETase and MHETase enzymes, which may open new avenue for bioremediation and removing plastics from the environment in a sustainable manner.

SCOPE OF THIS WORK

The work presented in herein is part of an ongoing project focusing on understanding the mechanisms of enzymatic hydrolysis of polyethylene terephthalate through various biochemical and microscopy approaches. The focus of the project includes further characterization of a polyethylene terephthalate hydrolase (referred to as PETase) and mono-(2-hydroxyethylene) terephthalate (MHET) hydrolase (referred to as MHETase), from *Ideonella sakaiensis* 201 F-6, initially described by Yoshida *et al.* in 2016[1]. The majority of the studies described in Chapters 1-5 have been conducted at the National Renewable Energy Laboratory (NREL), under the guidance of Dr. Gregg Beckham, who has conceived and authored the initial research article describing the crystal structure of PETase and a PETase protein mutant, with the latter having shown a higher efficiency of polyethylene terephthalate degradation[2]. The current work described here focuses on expanding our understanding of the mechanisms of substrate hydrolysis catalyzed by the PETase and MHETase enzymes, and potential methods for improving PET hydrolysis with these enzymes.

This thesis additionally describes methods of metabolite extraction from human plasma for -omics type approaches in Appendix A; and structure-function relationships in Archaeal transcriptional regulation in Appendix B. Results described in Appendix A & Appendix B were conducted under the guidance of Dr. Valérie Copié and Dr. Martin Lawrence, respectively.

CHAPTER ONE

SYNTHETIC POLYMER POLYETHYLENE TEREPHTHALATE AND ITS DEGRDADATION MECHANISMS

Introduction to Synthetic Polymers and (Poly)ethylene Terephthalate

A world without synthetic polymers (more widely known as plastics) is almost an unimaginable parallel universe. It is easy to overlook the importance of synthetic materials so integrated in our day to day lives – they have become almost invisible due to their wide ranging and common usage. How did synthetic polymers, products that were almost nonexistent 100 years ago, become so ubiquitous that we cannot imagine a material world without it? The ever-decreasing production costs, the versatility and the pliability (both mechanical and chemical) of man-made polymer materials certainly were responsible for their meteoric rise since their large-scale introduction during World War II[3]. Plastics have an undeniably significant impact on humanity's transcendence into a new era, called by many the "Plastic Age"[4], which we are currently witnessing. Their popularity of use has, however, come at an environmental cost: Because of their tendency to be utilized as single-use materials and their strong recalcitrance to biodegradation, enormous amounts of synthetic polymers in landfills and in Earth's ecosystems have been accumulating at a staggering rate[5]. Excluding the fibers, plastic production in the world totaled 348 million metric tons in 2017[6]. Yet, the amount of plastics that are recycled in the world is at an abysmal 9%[7,8]. As the increasing amount of synthetic polymers in the world is becoming more and more apparent, conservationists, scientists and world leaders

alike are declaring emergencies and trying to raise awareness, to prevent a possible global-scale pollution disaster due to plastic waste accumulation throughout our planet [9].

Polyethylene terephthalate (PET), first described and patented by Whinfield and Dickson in 1941[10], is one of the most commonly used synthetic thermoplastic polyesters. PET has since then seen applications especially in the textile, beverage, and packaging industries[11]. The highly desired aromatic thermoplastic polyester is synthesized from ethylene glycol (EG) and terephthalic acid, derived from the petrochemical aromatic precursor, xylene (Figure 1-1).

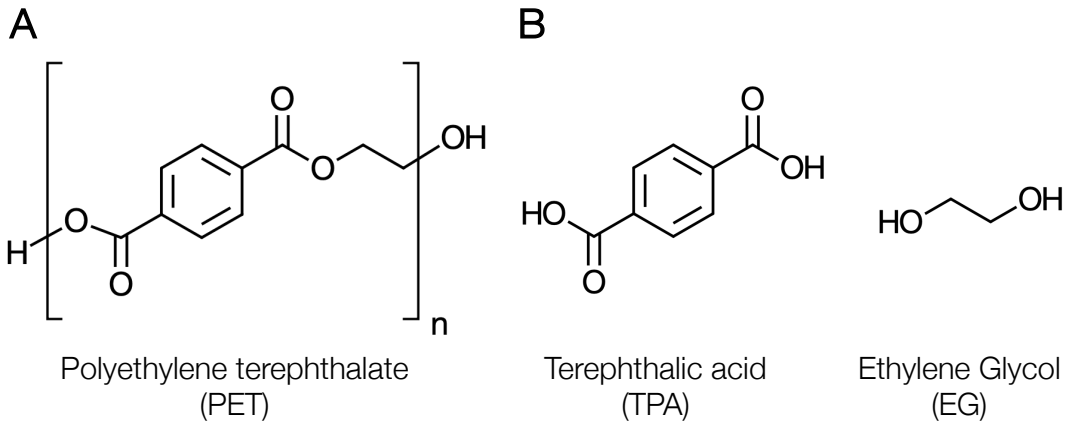


Figure 1-1: Chemical structures of (A) polyethylene terephthalate (PET), and its industrial precursors (B) terephthalic acid (TPA) and ethylene glycol (EG). Structures were drawn using ChemDoodle ver. 6.0[12].

Newly synthesized, virgin PET is a mostly amorphous material, whose properties can be selectively manipulated for desired applications[13,14]. PET is a highly hydrophobic substance with a glass transition temperature (T_g) of 72°C and can be heated above its T_g. Moreover, its polymer chains can easily be aligned to form either fibers made of

unidirectional PET molecules, or biaxially oriented molecules[15]. Both PET fibers and biaxially oriented PET are highly crystalline and are used commonly in the aforementioned industries. In crystalline PET, ester linkages that need to be accessible for PET depolymerization are masked, and chemical degradation of PET becomes a challenging process which requires a significantly higher monetary investment than the cost of producing PET[16–19]. Hence, PET and many other synthetic polymers share the same fate: They are either collected and stored in landfills, or, one way or another, end up in the environment, lingering for decades to come.



Figure 1-2: *Caretta caretta* turtle, entangled in a discarded plastic net. Photo by Eduardo Acevedo. Winner, Underwater Photographer of The Year 2019, in Marine Conservation category.

Synthetic Polymers as Environmental and Health Hazards

Geyer *et al.* reports that in 2015, 6.3 billion metric tons of single-use plastics were generated[7]. Besides their well-known physical damage, such as suffocation of marine

animals (Figure 1-2), plastics have been found to be toxic to flora and fauna of the ecosystems in which they reside[20]. Plastics now make up 61% of the worldwide beach litter[21], and all of the ocean basins have been found to be contaminated with plastics[22]. Furthermore, the existence of microplastics (debris smaller than 5mm) in the marine environments have revealed to be much more harmful than previously thought [7,23–25]. Even ecosystems that are seemingly untouched by humans are suffering from floating plastic patches[23]. In addition to the burden they impose on the world's ecosystems, their effects on human health are highly problematic: Americans currently consume somewhere between 74,000 to 121,000 microplastic particles in their diets and through inhalation[26]. Microplastics are now considered ubiquitous – so much that they are even present in commercial salt patches[27]. Even though microplastic ingestion has not been proven to be harmful in humans, proposed mechanisms of entry into cells, and effects of acute and chronic toxicity paint a dire picture[28].

While the recalcitrance of PET causes it to remain in the environment for prolonged periods, PET ,indeed, does degrade in the environment due to physical stresses, such as exposure to solar radiation, or mechanical stress[29,30]. This type of degradation, however, is harmful to the environment, as plastics that break down turn into microplastics which cause significant physical and chemical damage to marine fauna. Furthermore, upon exposure to UV, PET releases CH_4 and C_2H_5 aliphatic hydrocarbons, whose greenhouse effects are several times more potent than that of CO_2[31]. Considering the amount of plastics in the ocean and landfills, this recent finding indicates that plastics,

similar to food waste, may have a previously unrecognized potential to significantly impact anthropogenic greenhouse gas emission and climate change[32].

Realizing that the "plastics problem" requires immediate attention, the first official international policy, MARPOL 73/78, banning the disposal of plastics into the sea, was signed in 1988. However, the amount of plastic debris in the ocean has not decreased, despite the introduction of bans for plastic bags, plastic straws, microbeads and many single-use plastics since the early 1990's[33]. Additionally, policies replacing recalcitrant polymers with biodegradable polymers have been introduced in legislation, and while these policies may have some positive effects towards reduction of plastics being released into the environment, the planet's current state shows that these measures are not sufficient, and additional measures need to be taken.

Recycling of PET

Currently, the most common mode of recycling single use plastics is limited to mechanical recycling and energy production through thermal combustion[7]. As mentioned above, a very small percentage of plastics produced in the world is recycled, and even recycling has some drawbacks and environmental costs that need to be addressed.

PET itself is considered one of the best recyclable materials, along with aluminum[34]. In developed countries, economic incentives have managed to increase PET recycling rates up to 50%, but this number decreases all the way down to 1% in developing countries[35]. Mechanical recycling of PET introduces additional pitfalls: Material properties of mechanically recycled PET are not nearly as desirable as virgin

PET, making the reintroduction of mechanically recycled PET into global use more difficult[8,36]. Instead of mechanical recycling, chemical recycling of PET, i.e. conversion of the polymer into its precursors, is considered to be a more desirable approach. This stems from the fact that the precursors can be separated, purified, and re-used to produce virgin PET, which is in higher demand than its mechanically recycled counterpart. The cost of chemical recycling, however, is higher compared to that of mechanical recycling. Additionally, some chemical recycling methods employ the use of hazardous, almost opposite-of-eco-friendly catalysts, making the chemical recycling of PET a remaining challenge for the environment, and an active area of research[18,37].

Despite its current shortcomings, chemical recycling of raw materials and plastics has found a place at the heart of the concept of a circular economy[38], a system of practices aimed at "upcycling" already limited raw materials and precursors back into circulation via using sustainable approaches to regeneration of resources[39], thus preventing or minimizing the leakage of plastic waste into the environment.

Biodegradable and Biobased Plastics

Over time, there has been a major interest in plastics that are readily degraded to H_2O, CO_2, and biomass in the environment. Similarly, deriving polymers from biobased organisms such as plants and microbes have been of huge interest towards the goal of reducing humanity's dependence on petroleum, since most plastics manufactured today are petrochemical products. Derivation of PET building blocks from biobased sources has been slowly gaining industrial attention, and several studies have demonstrated viable

production of ethylene glycol (EG) and terephthalic acid (TPA) from plant-based sources have been made available in the past years[40,41].

It is important to note here that biobased plastics differ from biodegradable plastics. Indeed, biodegradable plastics, by definition, produce CO_2, H_2O and biomass upon degradation, while biobased plastics may not be readily biodegradable. Interestingly, many biodegradable plastics are polyesters, although their lifespans in the environment are shorter than that of their recalcitrant counterparts, such as PET. PET takes about 40 to 90 years to completely disintegrate, whereas biodegradable plastics such as PLA can full degrade within a year.[42] Many biodegradable plastics used today are aliphatic (some of which are also biobased) polymers, and except for polylactic acid (PLA), are aliphatic polyesters (Table 1-1, Figure 1-3), which are comprised of ester linkages that are more accessible to break down by hydrolase enzymes. However, some biobased plastics, such as PEF for example, cannot be readily biodegraded, and as such, their presence and impact on the environment are comparable to that of PET[43].

Table 1-1: Aliphatic, aromatic and semiaromatic polyesters that are produced from biobased sources.

Name of biodegradable plastic	**Abbreviation**	**Aliphatic/Aromatic?**	**Reference**
Polylactic acid	PLA	Aliphatic homopolyester	Mehta R., *et al.* (2005)[44]
Polyhydroxyalkanoates	PHA's	Aliphatic homopolyesters	Anjum, A., *et al.* (2016)[45]
Procaprolactone	PCL	Aliphatic homopolyester	Labet M. & Thielemans, W. (2009)[46]
Polybutylene succinate	PBS	Aliphatic homopolyester	Gigli, M., *et al.* (2016)[47]
Polybutylene succinate-*co*-adipate	PBSA	Aliphatic copolyester	Gigli, M., *et al.* (2016)[47]
Polybutylene adipate-*co*-terephthalate	PBAT	Aliphatic and aromatic copolyester	Ferreria, F. *et al.* (2019)[48]
Polyethylene 2,5-furandicarboxylate	PEF	Semiaromatic homopolyester	Han, H., *et al.* (2018)[43]

Enzymatic Depolymerization of PET

Enzymatic PET degradation efficiency is a multi-factor phenomenon. Alongside enzyme turnover rate, material properties such as crystallinity, glass transition temperature (T_g), and hydrophobicity are critical factors which affect efficient degradation, and as such, they have been of significant interest[49].

For enzymatic PET degradation to occur, several conditions need to be met: As PET is an insoluble substrate, microorganisms that are degrading the plastic need to adhere to the surface and secrete a PET hydrolase (and in some cases, more than just a PET hydrolase[1]) which can subsequently hydrolyze the ester linkages of PET, generating lower molecular weight oligomers and finally, monomers.

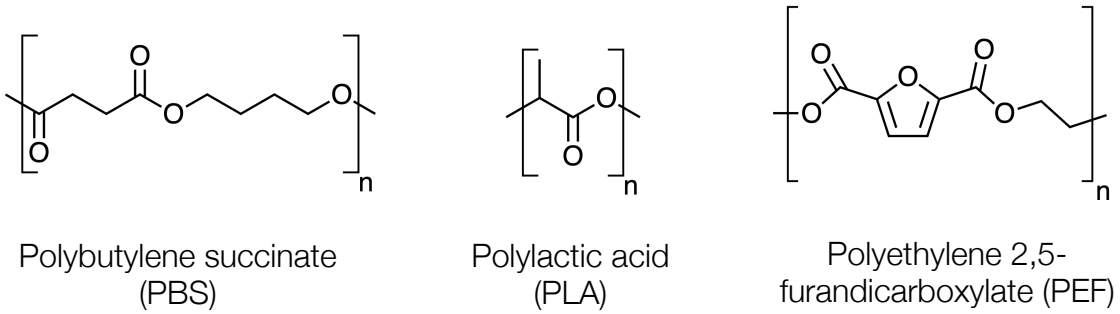


Figure 1-3: Notable biobased and biodegradable plastics mentioned in this work.

To date, several esterases (E.C. 3.1.1), few lipases (E.C. 3.1.1.3) and many cutinases (E.C. 3.1.1.74), isolated from bacteria and fungi, have been identified as having varying PET degradation capabilities [50–53]. Furthermore, regardless of their fungal or bacterial origin, crystal structures and homology models of known and putative PET hydrolases indicate that they all possess the well-known α/β hydrolase structural fold[54,55].

The canonical α/β hydrolase fold possesses a twisted, mixed β-sheet, wherein β-strands assume a $\beta_1\beta_2\beta_4\beta_3\beta_5\beta_6\beta_7\beta_8$ strand order, β_2 is antiparallel to the rest, and parallel β-strands are connected via right-handed α-helical turns (Figure 1-4). The amino acids involved in catalysis (i.e. the enzyme's catalytic triad) are located on surface accessible loops spanning the $\beta_5\beta_6$ and $\beta_7\beta_8$ strands, and on a C-terminal loop following β_8. The versatility of the fold also allows insertions of additional supersecondary structures and

domains within the loops separating the different β-strands. Lipases, for instance, exhibit mobile α-helical lid domains that are inserted into loops, creating a hydrophobic pocket surrounding the catalytic triad. α-helical domains can then partake in a hinge motion around the lipid-water interface, a process referred to as interfacial activation, allowing a more efficient hydrolysis of large hydrophobic fatty acid chains positioned in the micellar interface. Such interfacially activated lipases show enhanced activity towards their lipid substrates[56–58]. Interestingly, among the identified lipases that are capable of PET hydrolysis, the lid domains appear to be highly mobile, quite possibly allowing the entrance of the bulky PET functional groups into the active site of the lipase[60,61].

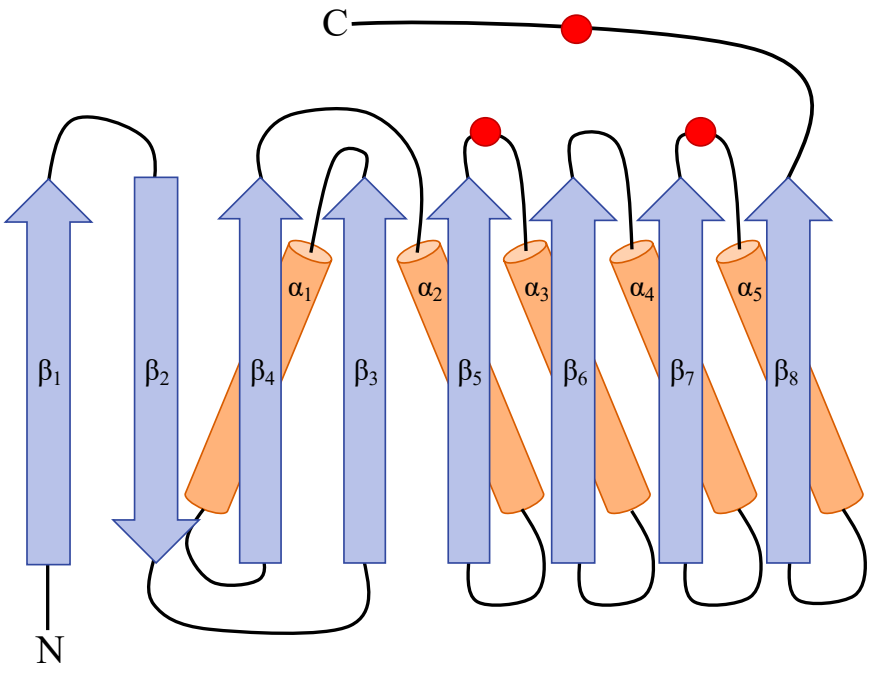


Figure 1-4: Topology diagram of a canonical α/β hydrolase fold, as described in SCOP2 database[59]. B-strands are depicted in blue arrows, whereas α-helices are depicted in orange. Red dots mark the loops in which the catalytic triad resides.

Ideally, enzymatic PET hydrolysis should take place right around the T_g of PET which is ~ 70°C. Several studies have identified thermostable cutinases capable of PET hydrolysis[62–64], but the organisms from which these proteins were identified are not capable of catabolizing aromatic building blocks such as TPA.

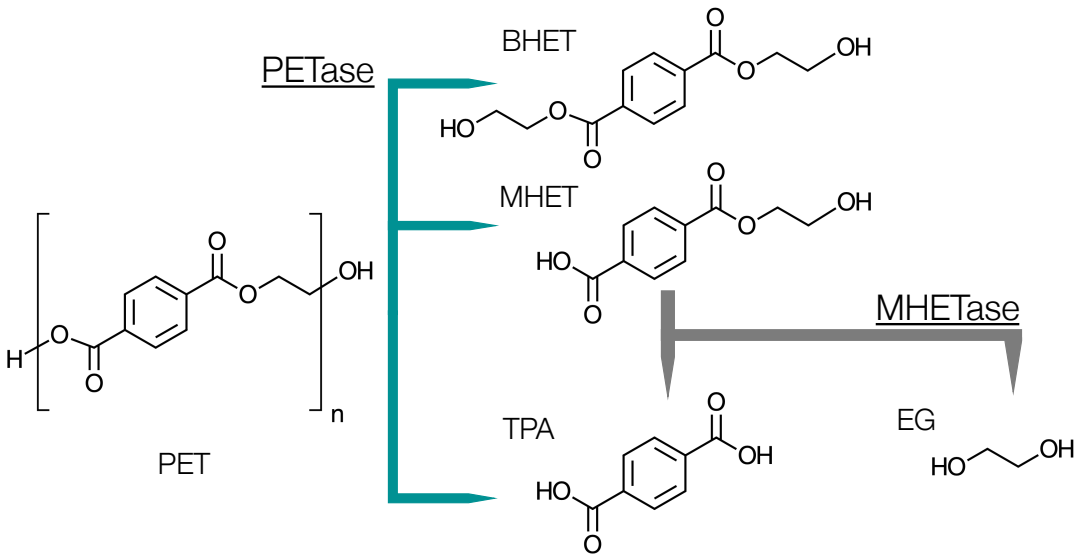


Figure 1-5: PET hydrolysis by PETase and MHETase. Hydrolysis of insoluble PET is carried out by PETase, resulting in the formation of the major hydrolysis product, MHET, and the byproducts BHET and TPA (shown in teal). The major hydrolysis product, MHET is then converted into TPA and EG by MHETase (shown in gray). Figure adopted from Austin *et al.*[2]

A Two-Enzyme System for PET Hydrolysis in *Ideonella sakaiensis*

Organisms such as *Comomonas sp.* E6 have been shown to degrade TPA through the known protocatechuate (PCA) pathway[65]. However, up until 2016, there were no reports of an organism who is capable of PET degradation and catabolism of PET monomers through any known pathway. In 2016, Yoshida *et al.* has identified *Ideonella*

sakaiensis 201-F6, an organism which can utilize PET and its monomer, TPA, as its major energy and carbon source[1]. In their report, Yoshida *et al.* showed that, when grown on PET film, *I. sakaiensis* increased expression of two proteins, one of which has high sequence identity to polyester degrading cutinases. This protein, aptly named PETase, is able to hydrolyze PET into mono(2-hydroxyethyl) terephthalate (MHET) as its major product, bis(2-hydroxyethyl) terephthalate (BHET) and TPA as its secondary products. The second protein, named MHETase, then hydrolyzes MHET into TPA and EG (Figure 1-5). Yoshida *et al.* additionally demonstrated that the genome of *I. sakaiensis* contains putative TPA degradation genes with high sequence identities to the TPA degradation clusters of *Comomonas sp.* E6. This further indicates that TPA and EG are indeed catabolized in *I. sakaiensis.*

PETase is an α/β Hydrolase

PETase from *I. sakaiensis* has been reported by Yoshida *et al.* as a few times more efficient at 30ºC compared to the cutinases reported in the literature[1]. While most reported cutinases have higher activities at 50-70ºC, PETase is not significantly active above 40ºC.

A number of reports have thus far crystallized and biochemically characterized PETase [2,66–69]. Table 1-2 provides PDB IDs of the structures and the resolution which the structures were solved. One of the first described cutinases that degrade PET, *Thermobifida fusca* cutinase (*Tf*Cut, PDB ID: 4CGI), has been shown to be a homolog of PETase.

Table 1-2: PETase structures deposited on Protein Data Bank (PDB) to date.

PDB ID	Resolution (Å)	Reference
6EQE	0.92	Austin *et al.* (2018)[2]
5XJH	1.54	Joo *et al.* (2018)[66]
5XG0	1.58	Han *et al.*(2017)[67]
6ANE	2.02	Fecker *et al.* (2018)[68]
5YFE	1.39	Liu *et al.* (2018)[69]
5XH3	1.3	Chen *et al.* (2018)[70]

In particular, Austin *et al.*[2] have identified structural features which were then exploited for a more efficient PET hydrolysis. The 30.2 kDa protein PETase (solved at 0.92Å resolution by single-wavelength anomalous dispersion (SAD)), much like its cutinase homologs, exhibits a canonical α/β hydrolase fold with a Ser160-Asp206-His237 catalytic triad sitting in the active site cleft (Figure 1-6, Figure 1-7 & Figure 1-8). Ser160 of the catalytic triad sits on the GX_1SX_2G consensus sequence, very common to esterases, known as the "lipase box". While Austin *et al.* have demonstrated that in the closest lipase and cutinase homologs of PETase, X_2 is usually a methionine in the lipase box, X_1 can be either a tryptophan or a histidine. In the closest PETase homolog with a known crystal structure, *Thermobifida fusca* cutinase (*Tf*Cut), X_1, is a histidine, and in PETase, X_1 (aka. position 159) is a tryptophan. Additionally, catalytic histidine (His237) of PETase is flanked by two serine residues (Ser236 and Ser238) both on N- and C-terminal ends, whereas *Tf*Cut catalytic serine is flanked by a serine and a phenylalanine. The two proteins that share 53% sequence identity[51] show a stark difference between the width of their active site clefts (8.46 Å in PETase vs. 2.98 Å in *Tf*Cut (Figure 1-7, Figure

1-8)), presumably because of the two aforementioned residues–Trp159 and Ser238, as Austin *et al.* underlines. Generation of a PETase double mutant, W159H/S238F, by homology modeling, presumably closes the gap of the active site cleft, thus making PETase more like *Tf*Cut.

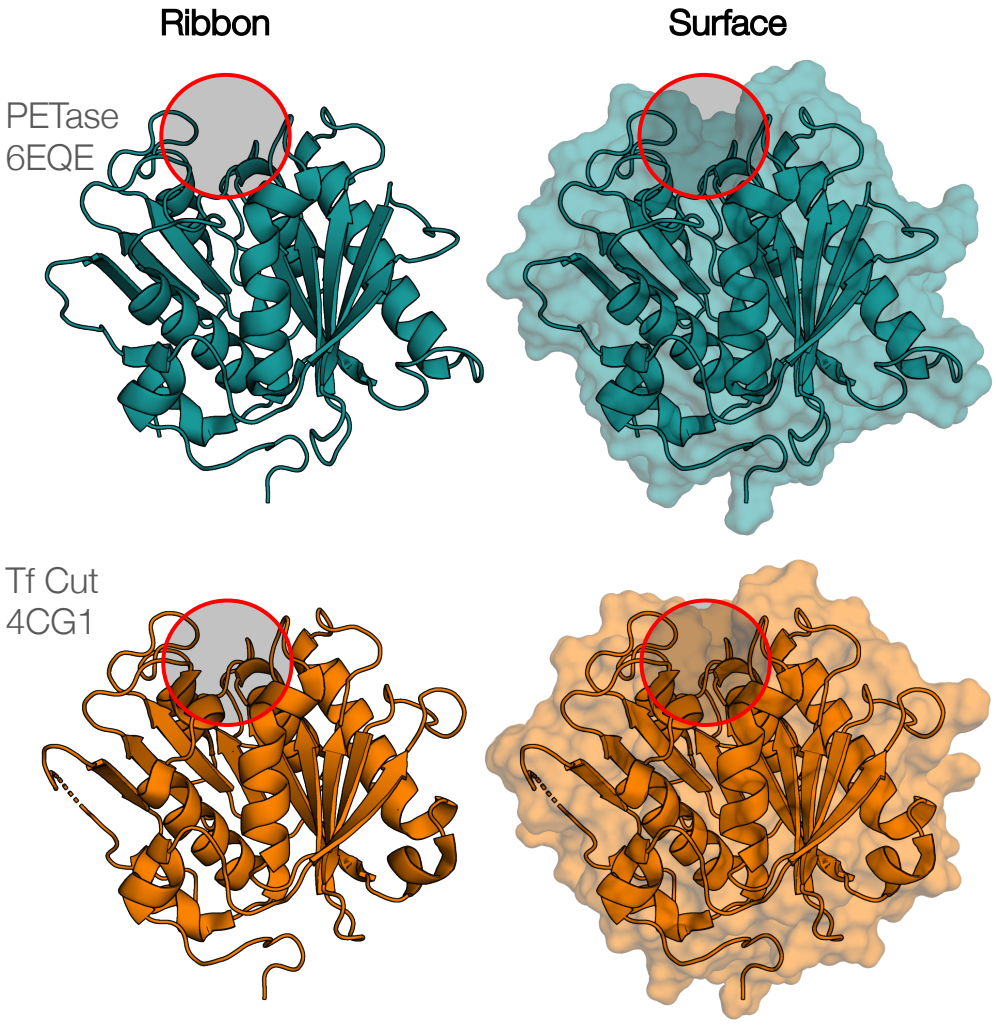


Figure 1-6: Structural comparison of PETase (cyan, PDB ID: 6EQE) and T. fusca cutinase (TfCut, orange, PDB ID: 4CGI), along with their surface representations. Active sites and the active site clefts of both proteins are circled red. Figure adapted from Austin et al. and ray traces were generated using PyMOL.

Incubation of wild type (WT) PETase and W159H/S238F double mutant with highly crystalline (14.8%) PET showed an unexpected result: After 96 hours, W159H/S238F has released more product (MHET and TPA) and has contributed more to the reduction of crystalline regions in PET, compared to WT PETase. The increase in enzyme effectiveness can be potentially contributed to an increased number of aromatic interactions Phe238 is facilitating during substrate binding.

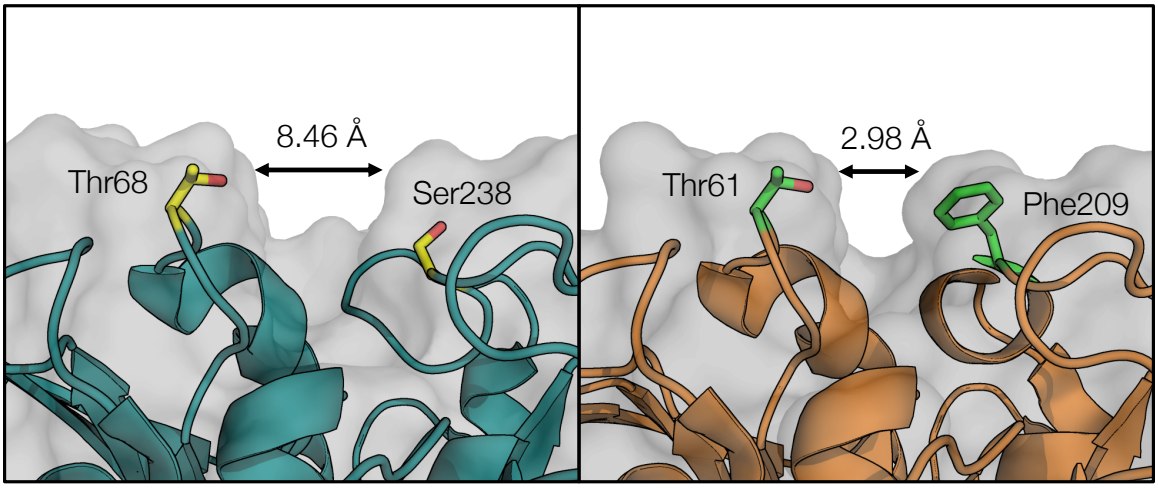


Figure 1-7: Active site clefts of PETase (left, teal, PDB ID: 6EQE) and *Tf*Cut (right, orange, PDB ID: 4CG1) respectively, compared. In PETase (left), the distance between Thr68 and Ser238 is 8.46 Å. In *Tf*Cut, the distance between equivalent positions of amino acids, Thr61 and Phe209 is 2.98 Å. Figure adapted from Austin et al.[2] and structures were generated using PyMOL.

To date, mechanism of action of PETase on PET has been studied by a few groups[66,67,71]. Cocrystallization of PETase with PET has not been successful thus far (factors such as crystal packing have been problematic), and molecular dynamics simulations, molecular docking and cocrystallization of molecules analogous to PETase substrates were utilized to elucidate the mechanism of PET hydrolysis[2,66–68]. It appears that PETase liberates MHET from PET via the conventional serine hydrolase mechanism.

There are still, however, questions that remain unanswered with respect to the how some of the residues rotate to accommodate PET during catalysis.

Interestingly, PETase has been demonstrated to work only on aromatic polyesters, and not aliphatic polyesters such as PBS and PLA[2] (Figure 1-1 & 1-3). Therefore, PETase can be considered an aromatic polyesterase, rather than just a PET hydrolase, whose specificity for aromatic polyesterases could prove advantageous for stepwise chemical recycling of aliphatic and aromatic polyesters in an industrial setting wherein aliphatic and aromatic postconsumer waste polyesters are present. Activity of PETase has been shown on PET and PEF, but its activity on copolymers such as PBAT (Table 1-1) has not yet been shown. Additionally, chemical basis for PETase specificity for aromatic polyesters is not yet elucidated.

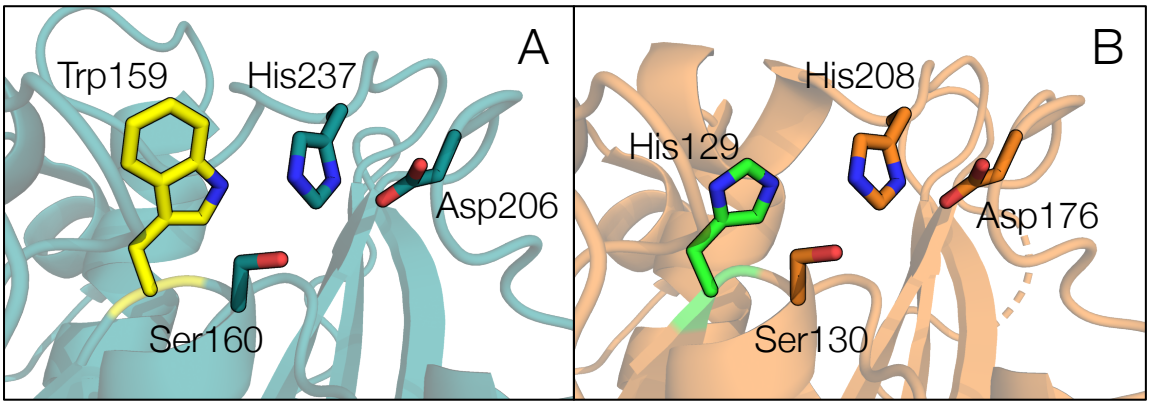


Figure 1-8: Active sites and the catalytic triads of PETase (teal) and *Tf*Cut (orange). A) PETase catalytic triad, S160, D206, H237 are represented as stick figures in teal and the S160 adjacent lipase box amino acid Trp159 is represented as stick figure in yellow. B) *Tf*Cut catalytic triad, S130, H208, D176 are represented as stick figures in orange and the S130 adjacent lipase box amino acid His129 is represented as stick figure in yellow. Figure adapted from Austin *et al.*[2] and generated using PyMOL[72].

Improved PET hydrolysis via engineering of PETase (as well as MHETase) can open the doors for industrial biobased plastics recycling in the near future. For this reason, understanding mechanisms of PET deconstruction processes is of significant importance.

Surface Degradation Capabilities of PETase

As mentioned above, material properties are major factors that contribute to efficient PET hydrolysis. PET hydrolysis, in general, requires enzyme access to the PET substrate, and therefore the PET polymer chains. Accessibility to the PET polymer chains is affected by several factors, with crystallinity of the polymer being one of the major contributors. In short, the more amorphous the PET, the easier is the enzymatic hydrolysis. For instance, hydrolysis at glass transition temperature, T_g, of a polymer (which can be amorphous or mostly amorphous with crystalline regions), can provide faster and more efficient hydrolysis, as the polymer chains at its T_g are at a glassy state, and thus move more freely. Since T_g of PET is 70-75°C, finding ways of efficient hydrolysis of PET by PETase is at temperatures greater than 30°C is very desirable. Additionally, the surface of PET is very hydrophobic; and its hydrolysis products, MHET and BHET have very little water solubility. Hydrophobic PET needs to be degraded by PETase, which has a very negative surface charge (and a pI of 9.6).

In a recent review, Kawai *et al.*[73] categorized enzymatic hydrolysis of PET by cutinases and lipases: PET modifying enzymes and PET hydrolases. The authors suggested that enzymes that do not degrade the inner block of PET substrate can be considered as PET surface-modifying enzymes, and not PET hydrolases. They

additionally argued that in order for a cutinase to be considered a "PET hydrolyzing" enzyme, it needs to possess the ability to degrade the inner block of the PET substrate and to cause visible surface erosion of the substrate. The enzymes that have been considered PET hydrolases (such as *Tf*Cut[51], LCC[74] and Cut190[52]) have all shown that they can degrade commercially available, low crystallinity PET (Table 2-1, lcPET, Kawai *et al.* refer to it as PET-GF).

On the other hand, several studies have investigated the degradation capability of PETase on PET substrates. In the report by Yoshida *et al.*[1], PETase is shown to degrade a synthesized, low crystallinity PET (1.9%). Moreover, PETase has been shown to degrade higher crystallinity, synthesized PET (~14%)[2]. Notably, degradation of higher crystallinity of PET by PETase has led to a decrease in crystallinity of the substrate. Conversely, treatment of lcPET (low crystallinity) by several different cutinases, such as *Humicola insolens* Cutinase (HiC), *Fusarium* solani Cutinase (FsC) and *Pseudomonas mendocina* (PmC) has shown an increase in crystallinity post treatment[62]. The majority of the presently reported studies have used high temperatures (50-70°C) for lcPET incubation, which are outside of the working temperature of PETase in its natural environment. Therefore, whether PETase can degrade commercially relevant substrates such as lcPET has not yet been demonstrated and is a relevant topic of research.

Moreover, the interplay and interactions between substrate and enzyme need to be elucidated in a clearer manner. As a lot of parallels can be drawn between enzymatic PET hydrolysis and enzymatic cellulose and chitin degradation, methods describing PETase

(and cutinase) action need to be expanded to visualization and modeling of enzyme activity on substrate surfaces[2,75].

Structural and Biochemical Features of MHETase

While PETase has received significant attention from several research groups, it is important to remember that in *I. sakaiensis,* PETase cannot fully convert PET into monomers that can be readily catabolized*,* and thus requires the presence of MHETase for completion of the PET degradation process. One of the striking features of MHETase is its specificity (and tight substrate binding) towards MHET (Figure 1-5). While it can hydrolyze MHET, MHETase cannot hydrolyze BHET. Yoshida *et al.*[1] and Palm *et al.*[76] have categorized MHETase as a protein that belongs to the tannase family (EC 3.1.1.20) which contains several tannases, feruloyl esterases (EC 3.1.1.73), and unknown proteins of bacterial and fungal origin (according to the ESTHER database). Tannases catalyze the hydrolysis of galloyl ester bonds in tannin. Feruloyl esterases, on the other hand, catalyze the conversion of ferulate esters in plant cell walls. Palm *et al.*[76] have reported the structure of MHETase which was solved using molecular replacement and using the *Aspergillus oryzae* feruloyl esterase *Ao*FaeB-2 (PDB ID: 6G21, 26% sequence identity). By structural comparison to *Ao*FaeB-2, Palm *et al.* have also engineered protein variants that can readily hydrolyze MHET and BHET.

Structure of MHETase by Palm *et al.*[76] sheds light on some of its biochemical features. There are, however, several questions that remained unanswered with regards to the mechanism of action of MHETase on MHET, as well as how it works at the molecular level in concert with PETase to promote near complete degradation of PET.

Significance of This Work

The negative environmental impact of petroleum-based products can be offset by the innovations that are being developed to manipulate and exploit plant-based biomass to create new, more environmental-friendly materials or to recycle chemicals (benzene, toluene, ethylene and xylene) for fuel production. While bioethanol and polylactic acid (PLA) produced from corn have gained significant attention, the use of agricultural land for ethanol-based materials and fuel can present its own set of problems, including increasing the cost of crops to feed humans and livestock, in addition to increased use of agricultural land [41,44]. While plants produce the majority of aromatic compounds that can be used as alternatives to petrochemicals, plants lack the ability to degrade and catabolize aromatic hydrocarbons, with the exception of plant enzymes that degrade aromatic amino acids. Consequently, bacteria and fungi are primarily responsible for recycling the carbon scaffolds present within highly stable aromatics compounds. For instance, there are microorganisms known to funnel lignin through several "upper" aromatic pathways to generate what are called "central metabolites" such as catechol (1,2-dihydroxybenzene) and protocatechuate (3,4-dihydroxybenzene). These aromatics can then undergo ring-modification and successive ring-cleavage reactions via *ortho*- and *meta*- cleavages to produce TCA cycle intermediates[77,78]. The strategies that are employed efficiently by microorganisms can be exploited to produce specific valuable small molecule intermediates. For instance, from lignocellulosic biomass alone, production of dicarboxylic acids such as fumaric acid, muconic acid and succinic acid can be achieved

and used in the formation of composite materials. In addition, muconic and succinic acids can be used to produce the high-commodity product: Adipic acid[79–83]. About 90% of world's adipic acid production goes to making Nylon 66, while the rest of the adipic acid produced is used in producing coatings and polyurethane materials. Muconic acid can also be employed as a precursor to terepthalic acid (TPA)[84]. Moreover, the unsaturated muconic and fumaric acids also serve as backbone precursors to unsaturated polyesters, which are also used in the production of composite materials[85]. It has been recently shown that *Pseudomonas putida* KT2440 can be engineered to produce several of these aromatic intermediates and can be used to make performance-advantaged materials, solely by growing on glucose[86], demonstrating the value of these enzyme engineering efforts to fulfill human needs for plastic materials, while minimizing their negative impacts on the environment and the planet.

Ideonella sakaiensis, with its ability to degrade PET, is thus an inspiration, and a good template to engineer new microorganisms that are both capable of degrading PET and catabolizing TPA efficiently, harnessing metabolic energy and facilitating aromatic hydrocarbon usage and recycling. These new biotechnological approaches promise to be extremely valuable to generate novel biobased and easily recyclable materials which will be critical to the development of a cost-effective circular economy that minimize environmental pollution.

Outline of Work with Polyesterases

This doctoral dissertation reports novel findings that enhance our knowledge of the diverse physical/mechanical and chemical mechanisms employed by PETase and MHETase enzymes to degrade PET. More specifically, Chapter 2 provides a detail account of the experimental approaches and analytical methods used to investigate the structural and catalytic properties of PETase and MHETase. Chapter 3 presents insights into the physical properties and mechanical ways used by PETase to preferentially hydrolyze PET regions of varying crystallinity and report several potential methods to improve PET hydrolysis by the wild type (WT) PETase enzyme, as well as a double protein mutant, W159H/S238F, of PETase, and colorimetric enzyme assays. Additionally, Chapter 3 also describes novel nanoscale experimental approaches to elucidate PET deconstruction at the molecular and nano-scale levels. Chapter 4 is focused on the biochemical, biophysical and structural characterizations of MHETase, and reports on bioinformatics, molecular dynamics, and structural approaches that have been employed to better understand MHETase enzymatic properties and biochemistry. Lastly, Chapter 5 synthesizes the work presented, and underlines various possible future directions that may be useful to further enhance our knowledge of PETase & MHETase-mediated plastic degradation processes.

CHAPTER TWO

MATERIALS AND METHODS

Introduction

This section describes the materials and experimental methods used to elucidate the mechanisms of PET hydrolysis by PETase and PETase kinetics (with results from these studies described in Chapter 3); In addition, this chapter reports on purification methods that have been employed to isolate MHETase protein mutants and associated enzyme activity assays, together with MHETase + PETase synergy assays (with results from these studies described in Chapter 4). Wild-type (WT) PETase enzyme, PETase W159H/S238F double mutant, PETase W185A single mutant, putative PETases from *Rhizobacter gummiphilus, Pseudomonas saudimassiliensis, and Pseudomonas bauzanensis;* and wild-type (WT) MHETase proteins used in this study were purified, lyophilized and sent by Mr. Harry Austin and Dr. Mark Allen in University of Portsmouth, United Kingdom.

Solid Substrate Assays

Table 2-1 tabulates the different types of substrates that have been used for the experiments. Two substrates ordered from Goodfellow, lcPET and boPET, were received as a 0.25 mm thick, clear sheet, and was cut into coupons with a hole punch. Coupons we have initially used for assays were circular with a ~6mm diameter. However, these coupons tended to float on the surface of the buffer solution, and therefore we have

switched to using a rectangular hole punch to ensure total submersion. Photographs of the coupons used in assays can be found in Chapter Three. Most of the other substrates were cut into coupons using a 4mm circular biopsy punch.

Masses of polymer coupons were measured prior to incubation with 50 nM enzyme (WT or mutant) in a 500 μL Eppendorf tube containing 50 mM bicine buffer (pH 9) or sodium phosphate buffer (at pH 7) for 96 hours at 30°C, 40°C, and 60°C. The reactions were quenched by incubating the reaction tubes at 80°C for 10 minutes and by adding an equal volume of phosphate buffer at pH 2.5. Coupons were then removed from the reaction mixture, and washed sequentially with 1% SDS, dH_2O, and 95% EtOH. Washed coupons were dried overnight in a vacuum oven at 60°C. Mass loss of the coupons was measured afterwards using an analytical scale (Mettler Toledo, Columbus, OH). Resulting supernatant was then centrifuged at 13,000 rpm on a benchtop centrifuge (Eppendorf Inc.) and used for analysis via high performance liquid chromatography (HPLC).

Addition of Surfactants to Solid Substrate Assays

Varying concentrations of Triton X-100 (0-25% v/v), and sodium dodecyl sulfate (SDS 0-0.04% w/v) were used to investigate the effects of surfactants on PET hydrolysis. Buffer conditions remained as described above.

Crystallinity Measurements of Polymer Samples

Thermal measurements of the polymer substrates were done using a TA Instruments Q-500 digital scanning calorimeter (DSC). Scans for the crystallinity

measurements were completed at a rate of 10°C/min from 60°C below the reported glass transition temperatures to 30°C above the melting temperature for each sample.

Table 2-1: Substrates used in this study.

Name of the substrate	**Abbreviation**	**Crystallinity**	**Obtained from**
Low-crystallinity polyethylene terephthalate	lcPET	~4%	Goodfellow, UK
Semicrystalline polyethylene terephthalate	semiPET	~12%	Goodfellow, UK
Biaxially-oriented polyethylene terephthalate	boPET	~40%	Goodfellow, UK
Fibrious polyethylene terephthalate	PET fiber	N/A	Goodfellow, UK
High-crystallinity polyethylene-terephthalate	sPET	~15%	Synthesized in-house
Polyethylene isophthalate	PEI	–	Synthesized in-house
Polyethylene 2,5 furandicarboxylate	PEF	~15%	Synthesized in-house
Polybutylene succinate	PBS	~42%	Synthesized in-house

Scanning Electron Microscopy

Dried coupons and substrates were examined by scanning electron microscopy (SEM). Samples were placed on an aluminum SEM stub with a carbon tape and sputter

coated with 9 nm of iridium. Conductive silver was applied to the coupons with heights greater than 200 μm to reduce charging of the plastic surface. Imaging was performed using FEG Quanta 400 under low vacuum (0.45 torr) with an accelerating voltage of 15 keV.

Ultramicrotomy of PET and PEF

Ultrathin (80-120 nm) sections of amorphous, semi-crystalline and biaxial PET substrates, as well as PEF, were cut using a Leica EM UC7 equipped with a diamond knife. Ultrathin sections were deposited on 300-mesh copper transmission electron microscopy (TEM) grids (SPI Supplies, West Chester, PA). TEM grids were in a reaction chamber. 5 μL, 50 nM enzyme containing 50 mM bicine buffer (pH 9) drops were deposited onto the TEM grid. A large reservoir of buffer solution was placed in the reaction chamber, which was then sealed with parafilm to prevent evaporation. TEM grid in the reaction chamber was incubated for 96 hours at 30°C. Afterwards, reactions were terminated by evaporating the drop at 30°C in the reaction chamber in open air. Grids were washed with single drops of dH_2O and 95% EtOH, respectively. Dried thin sections of grids were then stained with RuO_4 vapor for 6-10 hours (Figure 2-1). TEM micrographs were obtained using a Gatan UltraScan 1000 camera (Gatan, Pleasanton, CA) on a FEI Tecnai G20 Twin 200 kV LaB_6 transmission electron microscope using Digital Micrograph image capture software. Scanned micrographs were analyzed and thresholded using FIJI (NIH).

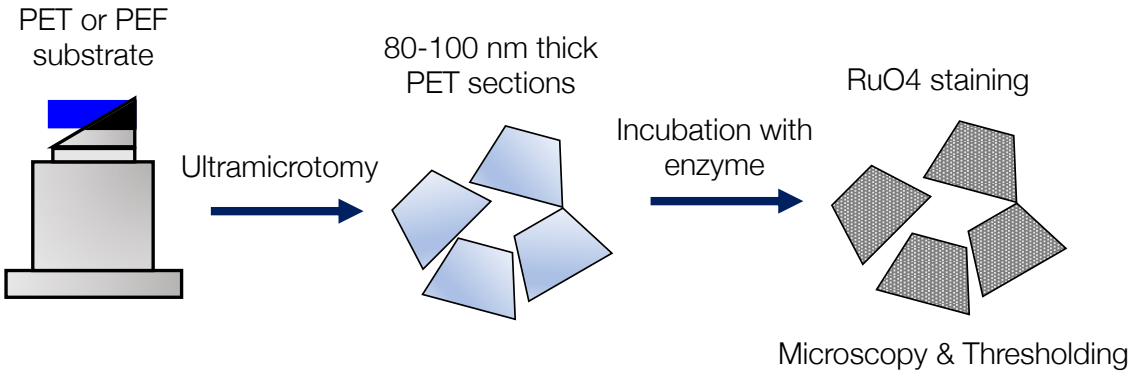


Figure 2-1: Workflow for digestion and analysis of thin-sectioned substrates.

Expression and Purification of Proteins

In our efforts to identify the critical amino acid residues and secondary structural features that are essential for PET, BHET and MHET hydrolysis, 1 PETase mutant and 14 MHETase mutants were engineered, and assessed for thermal stability and enzymatic activity. In addition to the protein mutants, genes coding for putative MHETase proteins from *Hydrogenaphaga sp.* and *Comomonas thiooxidans* were also cloned and the protein overexpressed in *E. coli.* C-terminal 6xHis-tagged gene constructs were generated by Mr. Graham Dominick (Manuscript in preparation) at the National Renewable Energy Laboratory. Brief descriptions of the PETase and MHETase protein mutants and potential MHETase enzymes from *Hydrogenaphaga sp. and C. thiooxidans* can be found in Table 2-2. Additionally, a detailed description of the methods for cloning and engineering of MHETase and PETase protein mutants is presented in the supplementary information associated in Chapter 4.

Small Scale Screening for Soluble Protein Expression

Gene constructs described in Table 2-2 and Table 2-3 were screened for soluble protein expression in *E. coli* using standard screening protocols [87]. Constructs that did not yield soluble protein expression in *E. coli* were tested for protein extraction from inclusion bodies and refolding protocols were used to try to regenerate active enzymes. Methods for optimization of protein expression included the use of several different T7 promoter- inducible *E. coli* strains, use of different growth media, assessment of the efficacy of autoinduction vs. IPTG induction[88], and cell growth culture temperature screening. Initial screening of mutants and putative MHETases was first undertaken using *E. coli* BL21 (DE3) (Stratagene), *E. coli* BL21 (DE3) RIL (Stratagene) and *E. coli* C41(DE3) (Lucigen) strains. Cells were then transformed according to manufacturer's instructions and plated on LB plates containing a final 100 μg/mL ampicillin or on LB plates containing a final 100 μg/mL ampicillin and a 34 μg/mL chloramphenicol. Single colonies selected from the LB plates were inoculated in 5 mL LB containing 100 μg/mL ampicillin and left to grow overnight at 37°C. 500 μl of these starter cultures were then added to 50 mL 2xTY medium containing 100 μg/mL ampicillin and were grown until OD_{600} reached 0.6-0.8. Isopropyl β-D-1-thigalactopyranoside (IPTG) was added to a final concentration of 0.25 to 1 mM to induce protein expression. Cell cultures were then grown for 6 hours at 16°C and harvested by centrifugation. Purification of the proteins of interest from resulting cell pellets were either carried out immediately, or harvested cell pellets were stored in -80°C until further use.

Table 2-2: Proteins expressed and purified in this study.

Protein	Plasmid	M.W. (kDa)	Description
01-PETlid	pCJ208	52.8	*I. sakaiensis* PETase with the lid region of MHETase. Trp185:Phe191 was deleted from PETase, Gly251:Thr472 from MHETase was inserted after Phe184 of PETase.
02-MHETnoLid	pCJ209	41.5	*I. sakaiensis* MHETase without its lid region. MHETase residues Gly251:Thr472 were deleted, PETase residues Trp185:Phe191 were added after Phe250 in MHETase.
03-MHETCC2AA	pCJ205	64.1	*I. sakaiensis* MHETase without of the disulfide bond adjacent to the active site (Cys224Ala/Cys529Ala)
04-MHETCC2WS	pCJ201	64.2	*I. sakaiensis* MHETase with active site-adjacent disulfide bond removed to made to look like the PETase motif. (Cys224Trp/Cys529Ser)
05-MHETCC2HF	pCJ204	64.2	*I. sakaiensis* MHETase with active site-adjacent disulfide bond removed to made to look like the double mutant PETase motif (Cys224His/Cys529Phe).
06-PETCC	pCJ202	30.2	*I. sakaiensis* PETase double mutant with an added disulfide bond adjacent to the active site (Ser238Cys/Trp159Cys).
07-MHETCat	pCJ196	64.1	*I. sakaiensis* MHETase active site serine mutated to alanine (Ser225Ala).
08-COMO sig	pCJ199	68.9	*Comomonas thiooxidans* putative MHETase with the signal peptide.
09-COMOnosig	pCJ203	60.3	*Comomonas thiooxidans* putative MHETase without the signal peptide.
10-PHAGAsig	pCJ207	62.3	*Hydrogenaphaga sp.* putative MHETase with its signal peptide.
11-PHAGAnosig	pCJ211	60.4	*Hydrogenaphaga sp.* putative MHETase without its signal peptide.
12-MHETnoLidCCwtPET	pCJ220	41.6	MHETase with no lid and its active site-adjacent disulfide bond changed to the WT PETase motif. MHETase residues Gly251:Thr472 were deleted, PETase residues Trp185:Phe191 were added after Phe250 in MHETase and Cys224Trp/Cys529Ser were added.

Table 2-3: Proteins expressed and purified in this study (continued).

Protein	Plasmid	M.W. (kDa)	Description
13-MHETnoLid CCdmPET	pCJ221	41.6	MHETase with no lid and its active site-adjacent disulfide bond changed to the WT PETase motif. MHETase residues Gly251:Thr472 were deleted, PETase residues Trp185:Phe191 were added after Phe250 in MHETase and Cys224His/Cys529Phe were added.
14-MHET131Gly	pCJ197	64.1	*I. sakaiensis* MHETase with Ser131Gly point mutation.
15-MHET495Ile	pCJ298	64.1	*I. sakaiensis* MHETase with Phe495Ile point mutation.
16-MHET6 CCFae	pCJ200	65.5	Addition of a 6th disulfide bond intoto *I. sakaiensis* MHETase from *Ao*FaeB-2. One point mutation (Ser136Cys), deletion of Trp75:His91 and addition of Tyr60:Phe89 from *Ao*FaeB-2 into MHETase.
17-MHETlip	pCJ206	64.1	*I. sakaiensis* MHETase point mutation in its kipase box (Glu226Thr).
18-MHET6 CCPET	pCJ217	64.2	Addition of a 6th disulfide bond into I. sakaiensis MHETase from *I. sakaiensis* PETase. Three point mutations (Ser136Cys/Gly489Cys/Ser530Cys),), deletion of Trp75:His91 in MHETase, and addition of Tyr60:Phe89 from AoFaeB-2 into MHETase.
19-MHET7CC	pCJ210	65.6	Addition of a 6th and 7th disulfide bond into I. sakaiensis MHETase from *I. sakaiensis* PETase. Three point mutations (Ser136Cys/Gly489Cys/Ser530Cys), deletion of Trp75:His91 in MHETase, and addition of Tyr60:Phe89 from AoFaeB-2 into MHETase.

Qualitative analysis of protein expression was accomplished using 1D SDS-PAGE. Among the screened proteins, several of them were found in high yields in the cell cultures following induction of protein synthesis using IPTG. Several other proteins, however, required the use of autoinduction medium for proper expression. Mutants of interest were at first used in kinetic assays, and therefore were not needed in high quantities. However, proteins that have shown promise in kinetic assays were further

improved for expression for future studies. Additionally, expression of putative MHETases from *Hydrogenaphaga sp.* and *C. thiooxidans* were optimized for future crystallographic studies, and their purification protocols are further described below.

Table 2-4: Results of small-scale expression and purification trials of proteins used in this study.

Protein	Residues	M.W. (kDa)	pI	Extinction coefficient ($M^{-1}*cm^{-1}$)	OD_{280} at 1 mg/ml	Soluble expression (mg/l)	SEC apparent M.W.
01-PETlid	505	52.8	8.97	89880	1.701	0.05	55 kDa
02-MHETnoLid	396	41.5	5.53	52370	1.259	0.1	45 kDa
03-MHETCC2AA	611	64.1	5.53	102330	1.596	No	
04-MHETCC2WS	611	64.2	5.53	107830	1.679	0.03	60 kDa
05-MHETCC2HF	611	64.2	5.59	102330	1.593	0.4	60 kDa
06-PETCC	290	30.2	9.48	33920	1.124	0.3	25 kDa
07-MHETCat	611	64.1	5.53	102330	1.595	1.2	64 kDa
08-COMOsig	650	68.9	6.79	102330	1.484	0.9	73 kDa
09-COMOnosig	575	60.3	5.53	96830	1.604	No	
10-PHAGAsig	592	62.3	5.31	92820	1.489	0.3	58 kDa
11-PHAGAnosig	573	60.4	5.23	92820	1.535	No	
12-MHETnolidCCwtPET	396	41.6	5.53	57870	1.395	0.1	45 kDa
13-MHETnolidCCdmPET	396	41.6	5.53	57870	1.389	0.1	45 kDa
14-MHET131Gly	611	64.1	5.53	102330	1.595	0.1	66 kDa
15-MHET495Ile	611	64.1	5.53	102330	1.596	0.04	66 kDa
16-MHET6CCFae	624	65.5	5.22	98320	1.5	0.01	66 kDa
17-MHETlip	611	64.1	5.61	102330	1.595	0.4	60 kDa
18-MHET6CCPET	611	64.2	5.53	102330	1.593	No	
19-MHET7CC	624	65.6	5.22	98320	1.499	0.2	60 kDa

Specific Approaches to Protein Expression and Protein Purification

01-PETlid, 04-MHETCC2WS For expression and purification of proteins PETlid and MHETCC2WS, *E. coli* BL21(DE3) RIL (Stratagene) cells were used. Cells were

transformed with either plasmids carrying the gene encoding PETlid or MHETCC2WS proteins (Table 2-2). Transformed cells were then inoculated in a starter culture of LB containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol until the cell cultures reached stationary phase at 30°C. The starter culture was inoculated at a 1000-fold dilution, into a 2xTY medium 100 μg/mL ampicillin and 34 μg/mL chloramphenicol and grown at 30°C until the OD_{600} reached 0.6-0.8. Protein expression was then induced by adding IPTG to a final concentration of 0.5 mM. Cells were grown for another 5 to 8 hours following IPTG induction, harvested by centrifugation, and stored at -80°C until use. Harvested cells were resuspended in a lysis buffer (300 mM NaCl, 20 mM Tris HCl, 10 mM imidazole, pH 8.0) and lysed using a bead beater. Lysate was clarified via centrifugation at 30,000 x g for 30-45 minutes. Clarified lysate was then applied to a 5 mL HisTrap Ni-NTA column on an ÄKTA Pure chromatography system (GE Healthcare). HisTrap HP column was washed with 5 column volumes of wash buffer (300 mM NaCl, 20 mM Tris-Cl pH 8.0, 10 mM imidazole), and 2 mL protein fractions were eluted with 3 column volumes of the elution buffer (100 mM NaCl, 20 mM Tris-Cl pH 8.0, 300 mM imidazole). Cell pellets, wash fractions and column eluents were analyzed for protein content using 1D SDS-PAGE with Coomassie staining and Western blots using 6xHis tag primary antibody (Invitrogen). Protein was then dialyzed with 100 mM NaCl, 20 mM Tris-Cl pH 7.0 overnight. Afterwards, protein was concentrated using a 10,000 Da molecular weight cut-off (MWCO) Amicon Ultra centrifugal filter and resulting filtrate was applied to a Sephacryl S-100 HR size exclusion column equilibrated

with 100 mM NaCl and 20 mM Tris-Cl pH 7.0. Protein concentrations were assessed using Bradford or BCA assays[89].

06-PETCC *E. coli* C41 DE3 (Lucigen) cells were used for expression and purification. Cells were transformed with the plasmid carrying the gene encoding the PETCC protein (Table 2-2). Transformed cells were then inoculated in a starter culture of LB containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol until the cell cultures reached stationary phase at 30°C. The starter culture was inoculated at a 1000-fold dilution, into a 2xTY medium 100 µg/mL ampicillin and 34 µg/mL chloramphenicol and grown at 30°C until the OD_{600} reached 0.6-0.8. Protein expression was then induced by adding IPTG to a final concentration of 0.5 mM. Cells were grown for another 5 to 8 hours following IPTG induction, harvested by centrifugation, and stored at -80°C until use. Harvested cells were resuspended in a lysis buffer (300 mM NaCl, 20 mM Tris HCl, 10 mM imidazole, pH 8.0) and lysed using a bead beater. Lysate was clarified via centrifugation at 30,000 x g for 30-45 minutes. Clarified lysate was then applied to a 5 mL HisTrap Ni-NTA column on an ÄKTA Pure chromatography system (GE Healthcare). HisTrap HP column was washed with 5 column volumes of wash buffer (300 mM NaCl, 20 mM Tris-Cl pH 8.0, 10 mM imidazole), and 2 mL protein fractions were eluted with 3 column volumes of the elution buffer (100 mM NaCl, 20 mM Tris-Cl pH 8.0, 300 mM imidazole). Cell pellets, wash fractions and column eluents were analyzed for protein content using 1D SDS-PAGE with Coomassie staining and Western blots using 6xHis tag primary antibody (Invitrogen). Protein was then dialyzed with 100 mM NaCl, 20 mM Tris-Cl pH 7.0 overnight. Afterwards, protein was concentrated using

a 10,000 Da molecular weight cut-off (MWCO) Amicon Ultra centrifugal filter and resulting filtrate was applied to a Sephacryl S-100 HR size exclusion column equilibrated with 100 mM NaCl and 20 mM Tris-Cl pH 7.0. Protein concentrations were assessed using Bradford or BCA assays[89].

02-MHETnoLid, 12-MHETnolidCCwtPET, 13-MHETnolidCCdmPET *E. coli* BL21(DE3) RIL (Stratagene) cells were used for expression and purification. Cells were transformed with each plasmid carrying the genes encoding MHETnoLid, MHETnolidCCwtPET and MHETnolidCCdmPET proteins (Table 2-2 and Table 2-3). Transformed cells were then inoculated in a starter culture of LB containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol until they reached stationary phase at 37°C. The starter culture was inoculated at a 1000-fold dilution, in a ZYP-5052 autoinduction medium containing 100 μg/mL ampicillin for 18-20 hours at 37°C. Cells were then harvested by centrifugation and stored at -80°C until use. Harvested cells were resuspended in a lysis buffer (300 mM NaCl, 20 mM Tris HCl, 10 mM imidazole, pH 8.0) and lysed using a bead beater. Lysate was clarified via centrifugation at 30,000 x g for 30-45 minutes. Clarified lysate was then applied to a 5 mL HisTrap Ni-NTA column on an ÄKTA Pure chromatography system (GE Healthcare). HisTrap HP column was washed with 5 column volumes of wash buffer (300 mM NaCl, 20 mM Tris-Cl pH 8.0, 10 mM imidazole), and 2 mL protein fractions were eluted with 3 column volumes of the elution buffer (100 mM NaCl, 20 mM Tris-Cl pH 8.0, 300 mM imidazole). Cell pellets, wash fractions and column eluents were analyzed for protein content using 1D SDS-PAGE with Coomassie staining and Western blots using 6xHis tag primary antibody

(Invitrogen). Protein was then dialyzed with 100 mM NaCl, 20 mM Tris-Cl pH 7.0 overnight. Afterwards, protein was concentrated using a 10,000 Da molecular weight cut-off (MWCO) Amicon Ultra centrifugal filter and resulting filtrate was applied to a Sephacryl S-100 HR size exclusion column equilibrated with 100 mM NaCl and 20 mM Tris-Cl pH 7.0. Protein concentrations were assessed using Bradford or BCA assays[89].

05-MHETCC2HF, 07-MHETCat, 14-MHET131Gly, 15-MHET495Ile, 17-MHETlip, 16-MHET6CCFae, 19-MHET7CC For expression and purification, OverExpress™ C41(DE3) competent cells were used. Cells were transformed with Transformed cells were then inoculated in a starter culture of LB containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol until they reached stationary phase at 37°C. The starter culture was inoculated at a 1000-fold dilution, in a ZYP-5052 autoinduction medium containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol for 18-20 hours at 30°C. Cells were then harvested by centrifugation and stored at -80°C until use. Harvested cells were resuspended in a lysis buffer (300 mM NaCl, 20 mM Tris HCl, 10 mM imidazole, pH 8.0) and lysed using a bead beater. Lysate was clarified via centrifugation at 30,000 x g for 30-45 minutes. Clarified lysate was then applied to a 5 mL HisTrap Ni-NTA column on an ÄKTA Pure chromatography system (GE Healthcare). HisTrap HP column was washed with 5 column volumes of wash buffer (300 mM NaCl, 20 mM Tris-Cl pH 8.0, 10 mM imidazole), and 2 mL protein fractions were eluted with 3 column volumes of the elution buffer (100 mM NaCl, 20 mM Tris-Cl pH 8.0, 300 mM imidazole). Cell pellets, wash fractions and column eluents were analyzed for protein content using 1D SDS-PAGE with Coomassie staining and Western

blots using 6xHis tag primary antibody (Invitrogen). Protein was then dialyzed with 100 mM NaCl, 20 mM Tris-Cl pH 7.0 overnight. Afterwards, protein was concentrated using a 10,000 Da molecular weight cut-off (MWCO) Amicon Ultra centrifugal filter and resulting filtrate was applied to a Sephacryl S-100 HR size exclusion column equilibrated with 100 mM NaCl and 20 mM Tris-Cl pH 7.0. Protein concentrations were assessed using Bradford or BCA assays[89].

Putative MHETase from *Comomonas thiooxidans* Two plasmids for the expression of the putative MHETase from *C. thiooxidans.* were initially constructed (pCJ199, pCJ203 (Table 2-2)). One of these plasmids, pCJ207, coded for the peptide signal sequence; pCJ211, on the other hand, did not. pCJ211 did not yield soluble protein expression in the trials, therefore pCJ207 was used to transform *E. coli* BL21(DE3) pLysS (Stratagene) cells for expression and purification. Transformed cells were then inoculated in a starter culture of LB containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol until they reached stationary phase at 37°C. The starter culture was inoculated at a 1000-fold dilution, in a ZYP-5052 autoinduction medium containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol for 18-20 hours at 30°C. Cells were then harvested by centrifugation and stored at -80°C until use. Harvested cells were resuspended in a lysis buffer (300 mM NaCl, 20 mM Tris HCl, 10 mM imidazole, pH 8.0) and lysed using a bead beater. Lysate was clarified via centrifugation at 30,000 x g for 30-45 minutes. Clarified lysate was then applied to a 5 mL HisTrap Ni-NTA column on an ÄKTA Pure chromatography system (GE Healthcare). HisTrap HP column was washed with 5 column volumes of wash buffer (300 mM NaCl, 20 mM Tris-Cl pH 8.0,

10 mM imidazole), and 2 mL protein fractions were eluted with 3 column volumes of the elution buffer (100 mM NaCl, 20 mM Tris-Cl pH 8.0, 300 mM imidazole). Cell pellets, wash fractions and column eluents were analyzed for protein content using 1D SDS-PAGE with Coomassie staining and Western blots using 6xHis tag primary antibody (Invitrogen). Protein was then dialyzed with 100 mM NaCl, 20 mM Tris-Cl pH 7.0 overnight. Afterwards, protein was concentrated using a 10,000 Da molecular weight cut-off (MWCO) Amicon Ultra centrifugal filter and resulting filtrate was applied to a Sephacryl S-100 HR size exclusion column equilibrated with 100 mM NaCl and 20 mM Tris-Cl pH 7.0. Protein concentrations were assessed using Bradford or BCA assays[89]

Putative MHETase from *Hydrogenaphaga sp.* Two plasmids for the expression of the putative MHETase from *Hydrogenaphaga sp.* were initially constructed (pCJ207, pCJ211 (Table 2-2)) to recombinantly express this potential MHETase enzyme. One of these plasmids pCJ203, coded for the peptide signal sequence; pCJ199, on the other hand, did not. pCJ203 did not yield soluble protein expression in the trials, therefore pCJ207 was used to transform *E. coli* BL2-CodonPlus (DE3) RIL (Stratagene) cells for expression and purification. Transformed cells were then inoculated in a starter culture of LB containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol until they reached stationary phase at 37°C. The starter culture was inoculated at a 1000-fold dilution, in a ZYP-5052 autoinduction medium containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol for 18-20 hours at 30°C. Cells were then harvested by centrifugation and stored at -80°C until use. Harvested cells were resuspended in a lysis buffer (300 mM NaCl, 20 mM Tris HCl, 10 mM imidazole, pH 8.0) and lysed using a bead beater. Lysate

was clarified via centrifugation at 30,000 x g for 30-45 minutes. Clarified lysate was then applied to a 5 mL HisTrap Ni-NTA column on an ÄKTA Pure chromatography system (GE Healthcare). HisTrap HP column was washed with 5 column volumes of wash buffer (300 mM NaCl, 20 mM Tris-Cl pH 8.0, 10 mM imidazole), and 2 mL protein fractions were eluted with 3 column volumes of the elution buffer (100 mM NaCl, 20 mM Tris-Cl pH 8.0, 300 mM imidazole). Cell pellets, wash fractions and column eluents were analyzed for protein content using 1D SDS-PAGE with Coomassie staining and Western blots using 6xHis tag primary antibody (Invitrogen). Protein was then dialyzed with 100 mM NaCl, 20 mM Tris-Cl pH 7.0 overnight. Afterwards, protein was concentrated using a 10,000 Da molecular weight cut-off (MWCO) Amicon Ultra centrifugal filter and resulting filtrate was applied to a Sephacryl S-100 HR size exclusion column equilibrated with 100 mM NaCl and 20 mM Tris-Cl pH 7.0. Protein concentrations were assessed using Bradford or BCA assays[89].

Kinetic Assays

PETase, MHETase and several MHETase protein mutants were incubated with various soluble substrates for determination of enzymatic activity. These substrates included: 4-nitrophenol acetate (*p*-NP acetate), 4-nitrophenol octanoate (*p*-NP octanoate), MHET, BHET and methyl ferulate. Kinetic assays for MHETase – MHET, MHETase – BHET, and MHETase – methyl ferulate are described below.

p-NP Acetate, *p*-NP Butyrate and *p*-NP Octanoate Assays

1 mM *p*-NP-acetate was incubated with enzyme in a 50 mM sodium phosphate buffer at pH 7 containing 10% DMSO at 30°C. Production of *p*-NP was monitored by recording absorbance at a wavelength of 405 nm over a time course of 10 minutes. Background absorbance was subtracted from the observed absorbance and resulting enzyme activity calculated using the GraphPad Prism 8 software.

Testing for Interfacial Activation of MHETase

MHETase was incubated with *p*-NP octanoate in a 10% Triton X-100 in 50 mM sodium phosphate buffer at pH 7 with 20-40% DMSO at 30°C.

MHETase & MHET Assays

Varying concentrations of MHET (1 μM to 20 μM) in 40 mM sodium phosphate buffer (pH 7), 80 mM NaCl, and 20% DMSO was incubated with 3nM protein (either WT or protein mutants) over the course of 30 minutes. Enzymatic reactions were stopped by adding an equal volume of 160 mM sodium phosphate buffer at pH 2.5 with 20% DMSO, and products of the reactions were analyzed by HPLC. Buffer conditions remained the same when testing MHETase activity for BHET and methyl ferulate degradation. Initial rates for each concentration were calculated using Graphpad Prism 8 software.

For turnover experiments with MHETase and MHET, 5nM protein in 40 mM sodium phosphate buffer (pH 7.5), 80mM NaCl, and 10% DMSO was incubated with 250 μM MHET and monitored for activity for 15 minutes. Enzymatic reactions were stopped

by adding an equal volume of 160 mM sodium phosphate buffer at pH 2.5 with 20% DMSO, and products of the reactions were analyzed by HPLC.

MHETase & BHET Assays

Varying concentrations of MHET (1 μM to 20 μM) in 40 mM sodium phosphate buffer (pH 7), 80 mM NaCl, and 20% DMSO was incubated with 3nM protein (either WT or protein mutants) over the course of 30 minutes for activity assays.

MHETase & Methyl Ferulate Assays

Varying concentrations of MHET (1 μM to 20 μM) in 40 mM sodium phosphate buffer (pH 7), 80 mM NaCl, and 20% DMSO was incubated with 3nM protein (either WT or protein mutants) over the course of 30 minutes.

PETase & BHET Assays

BHET (1 μl to 20 μl) in 50 mM sodium phosphate buffer (pH 7), 50 mM NaCl and 10% DMSO at 30°C was incubated with 3 nM WT PETase or the PETase W159H/S238F double mutant over a time course of thirty minutes. Enzymatic reactions were subsequently quenched by adding an equal volume of sodium phosphate buffer at pH 2.5 with 10% DMSO. Reaction tubes were centrifuged at 13,000 rpm for 10 minutes on a benchtop centrifuge, and later analyzed by HPLC.

PETase and MHETase Synergy

For PETase and MHETase synergy, effects of two independent factors, PETase load (mg/g PET) and MHETase load (mg/g PET) were explored as a function of mass loss of lcPET. For the initial explorative round, we have used a full factorial analysis. Each factor had enzyme load levels ranging from 0 mg/g PET to 3 mg/g PET with 0.25 mg/g PET increments in triplicates. Enzyme activity response was measured as lcPET mass loss following incubation at 30°C for 96 hours in 50 mM sodium phosphate buffer, pH 7. After the initial screening, optimal enzyme loading range of PETase and MHETase for improved mass loss was determined, a second round of optimization was conducted using surface response methodology. For the second round of optimization, enzyme loadings ranging from 0 to 0.375 mg/g PET for PETase and 0 to 0.75 mg/g PET for MHETase (in triplicates) were used. Enzyme activity response was measured as lcPET mass loss following incubation at 30°C for 96 hours in 50 mM sodium phosphate buffer, pH 7. For model generation, lcPET mass loss was used as a function of enzyme loading. Design Expert (Version 11.0, Stat-Ease Inc., USA) was used for response surface generation, evaluation of the model and the fit. The quadratic equation used to fit the model is as follows (Table 2-5 and Table 2-6): ln(mass loss) = -3.94 + 15.54352*A + 1.05674*B + 4.81714*AB – 29.66140*A2 – 2.53290*B2. The model was statistically analyzed to evaluate the analysis of variance (ANOVA) and fit (Table 2-7). The modeled surface was used to determine the optimum PETase and MHETase load for maximum response in terms of lcPET mass loss.

Table 2-5: Fit summaries of equations of various orders tested for response surface modeling.

Source	Sequential p-value	Lack of Fit p-value	Adjusted R^2	Predicted R^2
Linear	< 0.0001	0.0013	0.6128	0.5401
2FI	0.1219	0.0017	0.6368	0.5330
Quadratic	< 0.0001	0.6057	0.8383	0.7627
Cubic	0.3880	0.9862	0.8382	0.7433

Table 2-6: Equations of the quadratic model used to fit significant terms in the model.

Significant terms in the quadratic equation:	Intercept, A (PETase load), B (MHETase load), AB, A^2, B^2
Final equation in terms of coded factors (coefficient estimates):	ln(mass loss) = -1.70 + 1.17*A + 0.0226*B + 0.3387*AB – 1.04*A^2 - 0.3562*B^2
Final equation in terms of actual factors:	ln(mass loss) = -3.94 + 15.54352*A + 1.05674*B + 4.81714*AB – 29.66140*A^2 – 2.53290*B^2

Table 2-7: ANOVA for the quadratic model describing the PETase and MHETase synergy.

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	33.21	5	6.64	27.96	< 0.0001	significant
A-PETase load	24.54	1	24.54	103.30	< 0.0001	
B-MHETase load	0.0092	1	0.0092	0.0387	0.8459	
AB	1.38	1	1.38	5.80	0.0253	
A^2	6.52	1	6.52	27.47	< 0.0001	
B^2	0.7612	1	0.7612	3.20	0.0879	
Residual	4.99	21	0.2375			
Lack of Fit	0.4732	3	0.1577	0.6289	0.6057	not significant
Pure Error	4.51	18	0.2508			
Cor Total	38.20	26				

CHAPTER THREE

BIOCHEMICAL AND SURFACE CHARACTERIZATION OF HYDROLYSIS BY PETASE

Visualizing PETase Activity

It has been demonstrated that the enzyme efficiency of PET degradation can be altered by protein engineering, and promising results have paved the way for industrial applications of PETase (mentioned in Chapter Five). While the enzyme engineering aspect of this topic has gained significant traction, there is a need to better understand PETase's ability to hydrolyze surfaces in relation to surface morphology. In a recent review by Kawai *et al.*[73], PET hydrolysis by cutinases has been classified into two categories: Enzymes that modify PET surface but provide no surface erosion, and enzymes that hydrolyze PET and cause surface erosion. Today, low crystallinity PET (lcPET, Table 2-1) is used as an important substrate to investigate PET hydrolysis by cutinases, but the ability of PETase to degrade lcPET has not yet been demonstrated. Moreover, PETase's ability to hydrolyze PET substrates of various crystallinity and how the degradation process progresses still remain to be deciphered. To this end, we have assessed the capability of PET hydrolysis by PETase on various substrates, including lcPET. Furthermore, we have explored the nano-scale deconstruction mechanisms PETase might employ. We have also used several other putative PETases (described in Chapter Two; purified, lyophilized and sent to us by Dr. Mark Allen and Mr. Harry Austin at University of Portsmouth) for hydrolytic activity on PET. Nano-scale resolution

assays of lcPET and biaxially oriented PET (boPET, Table 2-1) were co-developed with Dr. Bryon Donohoe at National Renewable Energy. Funding for the surface characterization work done in NREL's Biomass Surface Characterization Laboratory (BSCL) was provided by Dr. Gregg Beckham's BETO and LDRD grants.

Preferential Hydrolysis by PETase

PETase from *I. sakaiensis* has the ability to degrade synthesized, low crystallinity (1.9%) PET (Synthesized by Yoshida and colleagues), as well as high crystallinity PET (sPET, ~14%)[1,2]. To that end, we have explored PETase's potential hydrolysis capability of lcPET. Firstly, incubation of lcPET with WT PETase and W159H/S238F at 30°C for 96 hours has shown morphological changes to lcPET, which are visible to the naked eye. These changes were then further investigated using SEM.

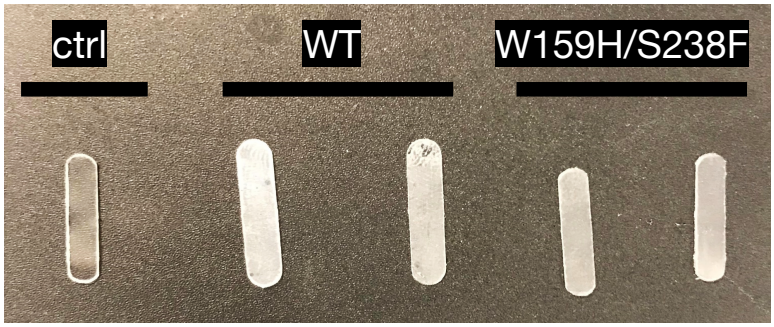


Figure 3-1: Changes visible to the naked eye in the surface of lcPET by treatment at 30°C for 96 hours with no enzyme (buffer only) control, wild type PETase, and the double mutant PETase. ctrl: no enzyme control, WT: Wild type PETase, W159H/S238F: Double mutant PETase.

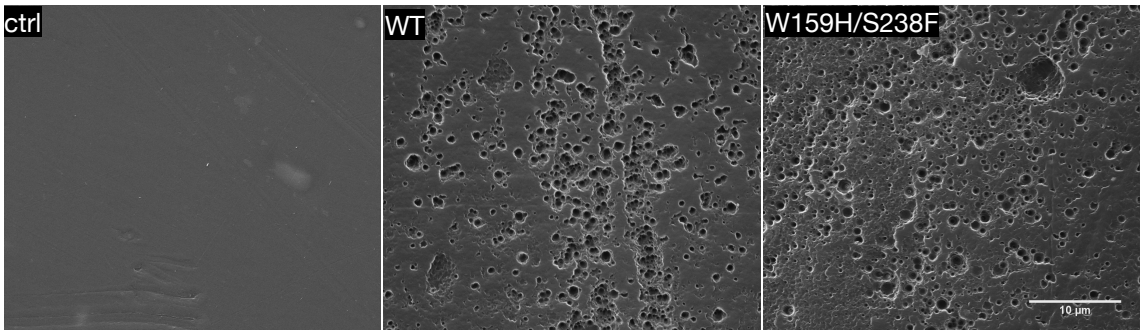


Figure 3-2: Surface pitting observed in lcPET by treatment at 30°C for 96 hours with no enzyme (buffer only) control (left panel, denoted ctrl), wild type PETase (middle panel, denoted WT), and the double mutant PETase (right panel, denoted W159H/S238F). Scale bar: 10 µm

SEM images of lcPET revealed that PETase creates surface pitting (otherwise referred to as surface erosion by Kawai et al., ((Figure 3-2). The surface pitting observed in lcPET are similar to what has been observed by Yoshida et al. and Austin et al. Visible pitting is accompanied by mass loss of lcPET (Figure 3-3). As expected, lcPET coupons that are treated with the W159H/S238F PETase mutant enzyme lose significantly more mass over the course of 96 hours, compared to WT PETase incubated lcPET and the no enzyme present control (Figure 3-3). Surface pitting created by PETase in the coupons provides the basis for the morphological changes that are observable to the naked eye: A clear coupon becomes opaque from the light scattering caused by the pitted surface, similar to frosted glass. This “frosting effect” also helps with quick and qualitative assessment of surface degradation.

Upon closer look at the replicates of lcPET degradation, we have additionally noticed that PETase hydrolysis patterns differ from one side to the other of the same

coupon that is completely submerged in buffer containing the enzyme (Figure 3-4 A, B D & E). Surface pitting in one side of the coupons are observed to be preferred.

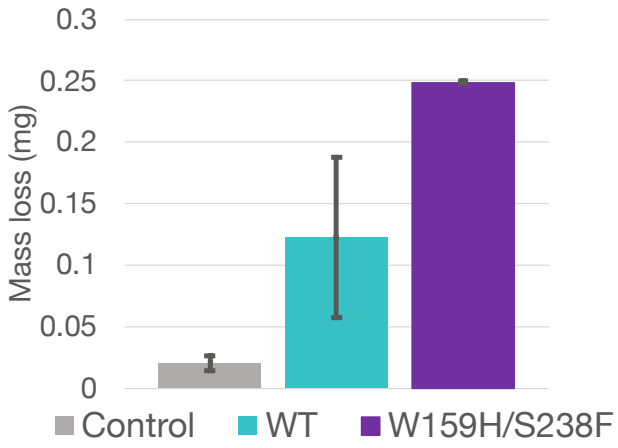


Figure 3-3: Mass loss of lcPET coupons upon incubation with buffer only (no enzyme) control (gray), WT PETase (teal) and W159H/S238F double mutant (purple). Values shown are the mean of three replicates and error bars indicate the standard error.

compared to the other side, and this preference is apparent even to the naked eye. Furthermore, SEM micrographs revealed that different sides of the same lcPET coupon also exhibit different hydrolysis patterns of hydrolysis by PETase (Figure 3-3; Figure 3-4 C, D & E), while maintaining the visual cue of “frosting effect” mentioned previously.

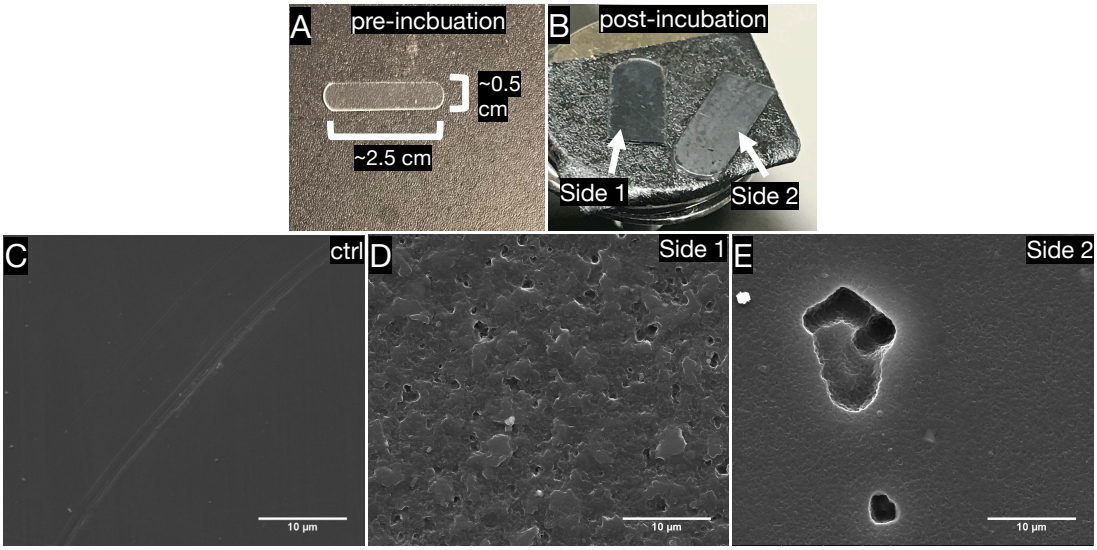


Figure 3-4: Preferential hydrolysis patterns by WT PETase. **A)** Dimensions of lcPET coupon cut with a hole punch. **B)** Same lcPET coupon cut in two with its different sides on an aluminum SEM stub. Side 1 and side 2 are marked with arrows. **C)** SEM micrograph of the surface of lcPET incubated with no enzyme (buffer only) control. **D)** SEM micrograph of the side 1 of lcPET coupon after incubation with WT PETase. **E)** SEM micrograph of the side 2 of lcPET incubated with WT PETase. Scale bar: 10 µm

It should be noted that Figures 3-2 and 3-4 represent both feature SEM micrographs of the same substrate (lcPET) which were conducted in technical replicates. Furthermore, the coupons were cut from the same lcPET film, prior to digestion. Stemming from this observation of varying patterns of hydrolysis, we incubated several more lcPET coupons for further inspection under SEM, with the intention to categorize more, if any, types of PET deconstruction processes. Out of 10 lcPET coupons and observing each of their sides following PETase hydrolysis, we have noticed two types of surface morphology: Cratered surfaces (similar to Figure 3-2 and Figure 3-4 E) and craggy, uneven surfaces (Figure 3-4 D). While 3 out of 10 of the lcPET coupons exhibited only one type of surface abrasion on both sides, the other lcPET coupons exhibited the same pattern that we initially observed, i.e. two sides of the same coupon

with differing hydrolysis patterns. We initially suspected different hydrolysis rates by PETase from each surface; unfortunately, quantification of product formation on one side of the coupon compared to the other is not possible via HPLC, on account of the two sides being present in the same reaction tube. Nevertheless, we never observed coupons with varying surface abrasions on the same side, indicating that mechanical features which are related to how the lcPET film is formed may be at play. It is possible that during the production process, one side of the mold-injected thin film of amorphous PET participate in interactions with the mold that may facilitate ordering of the polymer chains, whereas the other side is exposed to air.

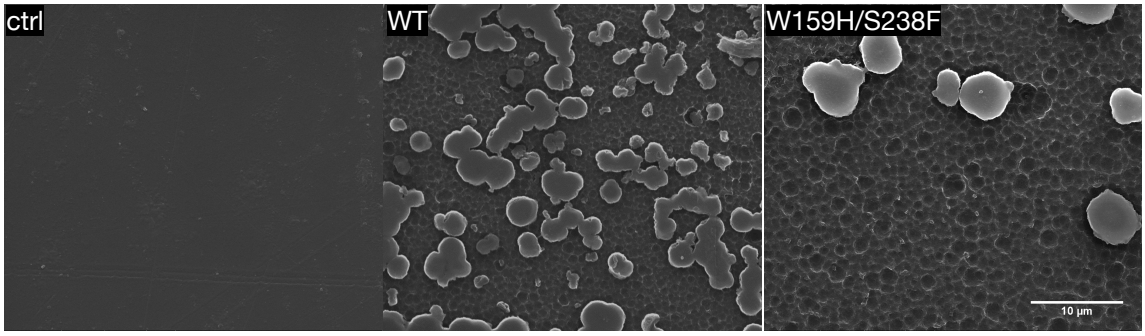


Figure 3-5: Incubation of sPET with no enzyme (buffer only) control (left panel, denoted ctrl), wild type PETase (middle panel, denoted WT), and the double mutant PETase (right panel, denoted W159H/S238F). Scale bar: 10 µm

Hydrolysis of Various Substrates by PETase

In addition to lcPET, we investigated the degradation capabilities of PETase on other substrates, including the industrially relevant biaxially oriented PET (boPET, used in beverages), and the high crystallinity, in-house synthesized PET (sPET). Incubation

with sPET yielded results consistent with the published report of Austin et al. (Figure 3-5). The “islands” seen in these SEM images prompted the idea that these structures might indeed be spherulites, i.e. lamellar morphology regions described in Chapter One.

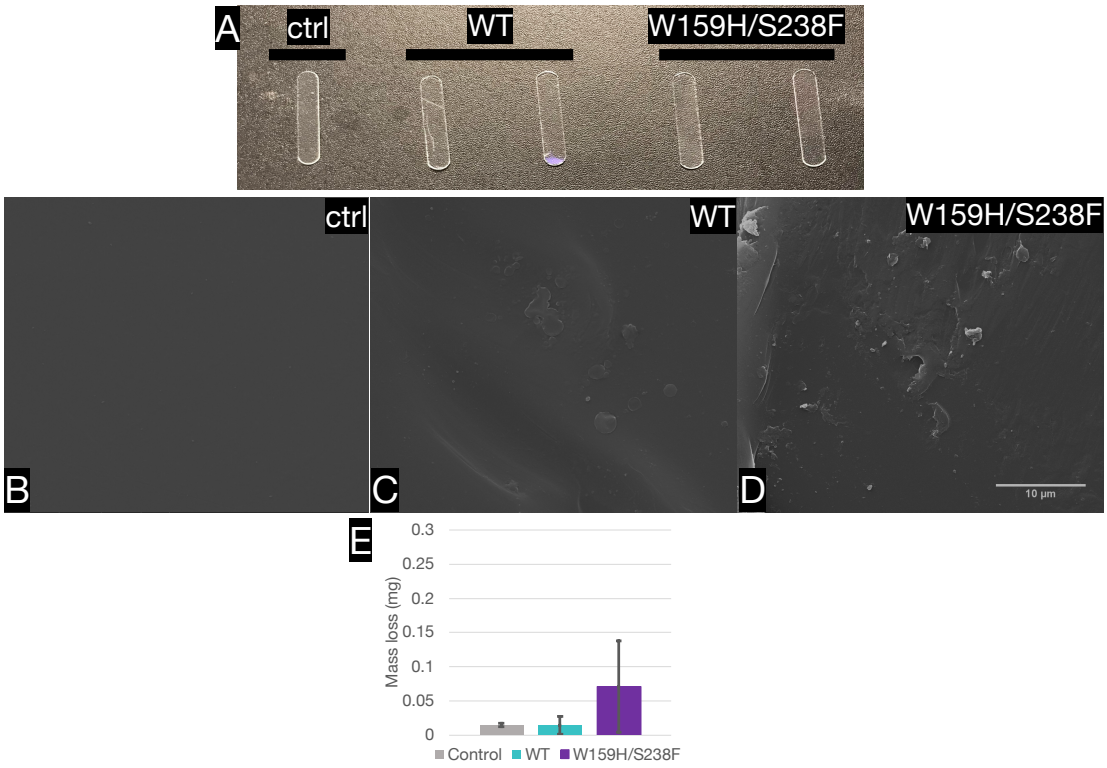


Figure 3-6: Incubation of boPET with WT PETase and W159H/S238F. A) Photograph of boPET coupons after treatment with buffer only control (ctrl), wild type PETase (WT), and double mutant PETase (W159H/S238F). B) SEM micrograph of boPET surface with no enzyme (buffer only) control. C) SEM micrograph of boPET with WT PETase. D) SEM micrograph of boPET surface with W159H/S238F. E) Mass loss of boPET after 4 days. Gray: Buffer-only control; Teal: WT PETase, Purple: W159H/S238F. All incubations took place at 30°C for 96 days, in 50 mM sodium phosphate buffer, pH 8. Scale bar (for B, C & D): 10 µm

Incubating boPET with WT PETase or the W159H/S238F PETase double mutant yielded no visible surface modifications, no observable surface hydrolysis (i.e. pitting)

(Figure 3-6), and no measurable mass loss. However, in later experiments with boPET and ~30% crystalline PET, we observed product release by HPLC, albeit below the quantifiable levels, even though both WT PETase and the W159H/S238F PETase double mutant suffers from not being able to readily access the biaxially oriented polymer chains. Numerous solid surface deconstruction strategies by functionally distinct cellulases have been reported in the literature[75,90–92]. While we suspected that there was not an intricate, unidirectional, exo-acting mechanism (such as that of Cel7a from *Trichoderma reesei*) at work in the case of PETase, we nevertheless examined PETase's ability to degrade plastic fibers, that are used prominently by the carpeting and textile industry. We found that PETase was able to degrade plastic fibers after four days with relative ease, with a tip-sharpening like effect on the fiber, although abrasions on the surface of the fiber was also observed (Figure 3-6). We did not note any discernable differences in deconstruction mechanisms between different replicate experiments on the fibers.

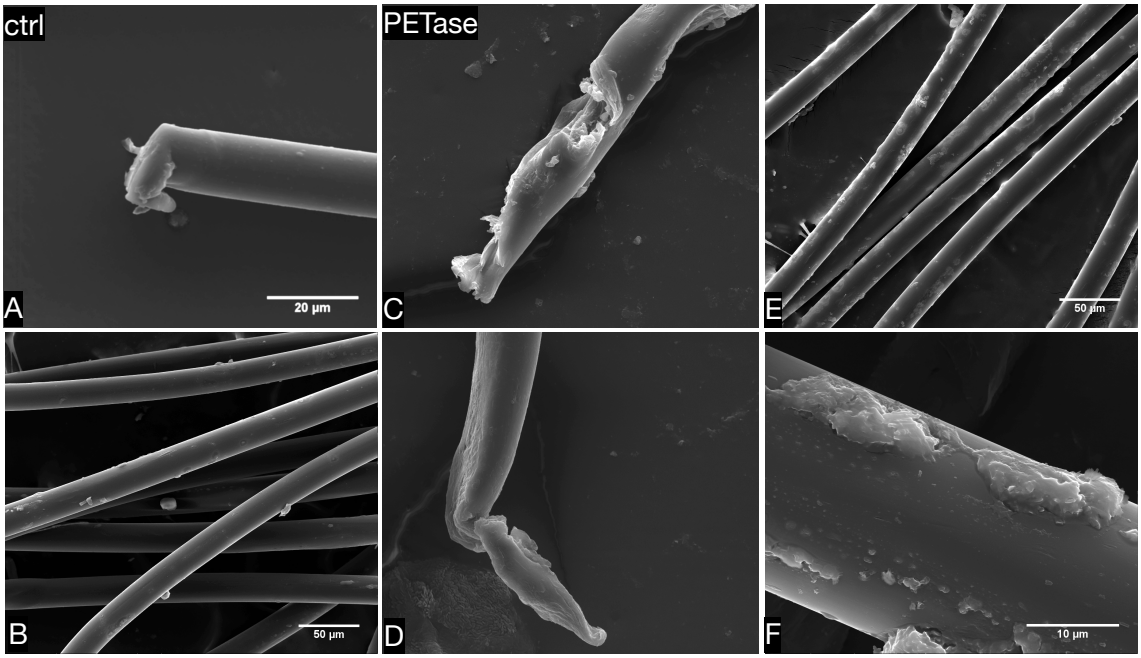


Figure 3-7: Hydrolysis of PET fibers with buffer only (no enzyme) control (denoted ctrl, A&B) and WT PETase (C, D, E & F). Scale bar for A, C &D: 20µm. Scale bar for B & E: 50 µm, Scale bar for F: 10 µm.

Unraveling Nano-scale Features of PET Deconstruction Mechanisms by PETase

While we know that crystallinity is a significant factor in determining PET hydrolysis, we do not, however, fully understand the effect of crystallinity and PET morphology at a nanoscale on the PETase enzyme function and efficiency. To that end, we attempted mounting enzyme-treated and untreated PET fibers on TEM grids. This, however, proved unsuccessful as PET fibers were found to be unfortunately too thick for TEM analysis. Additionally, staining of PET for TEM poses a problem: With regards to staining polymers for TEM analysis, OsO_4 is usually the oxidizer of choice. However, the oxidation capabilities of OsO_4 are inadequate for aromatic polyesters. In a report by

Haubruge and colleagues, RuO_4 vapor-stained thin sections of melt-crystallized PET were inspected using TEM to assess the lamellar thickness of samples with varying crystallinity[93]. RuO_4 can easily stain the amorphous regions of the polymer due to the fine RuO_2 particle precipitation in the amorphous regions. Meanwhile, in the more ordered regions, lamellae of the polymer cannot be fully accessed by the RuO_2 particles, resulting in the nano-scale resolution of the lamellae (Figure 3-8 and Figure 3-9). This allows for the inspection of the crystalline morphology of PET by TEM. Based on this observation, we cut thin sections (90 to 120 nm) of lcPET, boPET and optimized staining with RuO_4 to examine their morphology before and after PETase digestion.

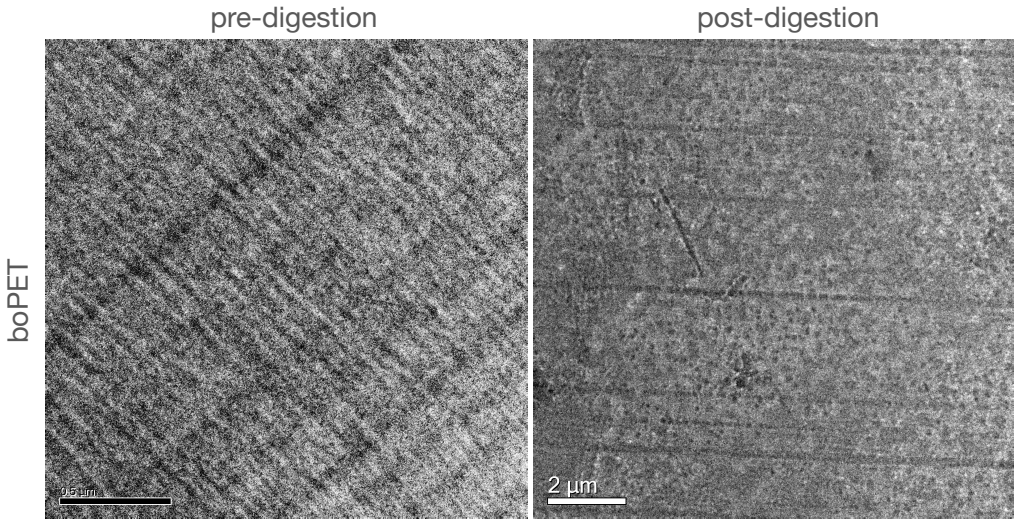


Figure 3-8: Representative TEM micrographs of RuO_4 stained boPET, pre and post-digestion by PETase. pre-digestion scale bar: 0.5µm, post-digestion scale bar: 2µm

Firstly, we evaluated the amorphous morphology of lcPET prior to PETase hydrolysis. Before incubation of lcPET with enzyme, we estimated the lamellar widths of lcPET to be 2.0 nm on average, with a percent order of ~18 % (Figure 3-9, Table 3-1).

Following enzyme digestion, we estimated that size of the average lamellar width to be larger, and percent order of the lcPET thin section to be increased to ~24%.

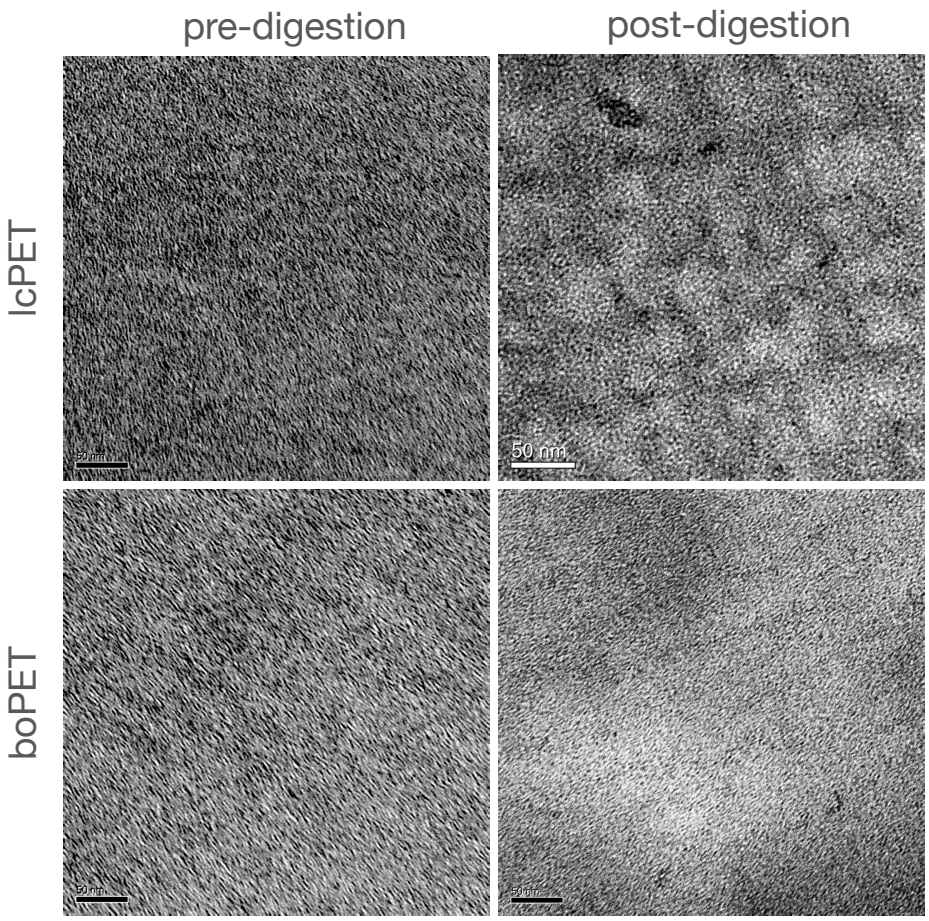


Figure 3-9: TEM micrographs of lcPET and bo PET pre and post-digestion. Lamellar widths and percent order for both substrates (Table 3-1 & Table 3-2) are calculated from binary images (described in Chapter 2). Scale bar: 50 nm.

Next, we moved to investigating the morphological features of boPET. Estimated lamellae widths prior to enzyme digestion averaged ~2.5 nm, compared to the 2.0 nm of lcPET, consistent with the fact that boPET is more crystalline than lcPET. Additionally, the percent order of biaxial PET was estimated to be at ~45%, which is surprisingly close to its crystallinity percentage, although order and crystallinity should not be confused

with each other (Table 3-2), as order, in this case, is a local estimation of the crystallinity, and may not apply to the bulk of the polymer. Although PETase does not create surface erosion on boPET (Figure 3-5), product hydrolysis is observed by HPLC, and thus it can be considered that PETase is indeed participating in some degree of boPET hydrolysis. While we cannot observe the patterns of hydrolysis with respect to order (and crystallinity) via SEM, our TEM micrographs of boPET following digestion with PETase reveal that the boPET thin sections are indeed going through some morphological change (Figure 3-8, Figure 3-9). Our initial calculations of lamellar width and percent order for boPET post-digestion yielded some interesting results, including a decrease in lamellar width from an average of 2.5 nm to 2.4 nm, and a percent order reduced to 39 ± 2%. While these results are consistent with the Austin et al. report which indicated a reduction in crystalline domain size following PETase digestion, substrates used in this study differ (lcPET & boPET vs. sPET).

Table 3-1: Calculated lamellar widths of lcPET pre and post PETase digestion. Std dev.: Standard deviation.

lcPET lamellar thickness (l_c)

	Pre-digestion Lamellar Width	**Post-digestion** Lamellar Width
Range	1.1-3.5 nm	1.9-9.3 nm
Average	2.0 nm	3.5 nm
Std. dev.	0.5 nm	1.2 nm
	Percent order	Percent order
	18 ± 2%	24 ± 6%

Table 3-1: Calculated lamellar widths of lcPET pre and post PETase digestion. Std. dev.: Standard deviation.

boPET lamellar thickness (l_c)

	Pre-digestion Lamellar Width	**Post-digestion** Lamellar Width
Range	1.4-5.0 nm	1.2-3.1 nm
Average	2.5 nm	2.4nm
Std. dev.	0.7 nm	1.2 nm
	Percent order	Percent order
	45 ± 4%	39 ± 2%

Effect of Temperature and Various Additives on PET Degradation

Temperature We wished to further explore how and in which ways the W159H/S238F PETase double mutant performs better than WT PETase. Firstly, we looked at the effect of temperature, and for this purpose, lcPET was incubated for 4 days with WT PETase and W159H/S238F under varying temperature conditions ranging from 20°C to 40°C. After 4 days, lcPET mass loss by WT PETase was highest at 30°C and 35°C, with a maximum mass loss of 0.1345 ± 0.0229 mg at 30°C and 0.1452 ± 0.0412 mg at 35°C (Figure 3-10). While W159H/S238F performs slightly better at 30°C than WT PEtase, it, interestingly, displayed significantly more mass loss at 40°C, and a temperature at which WT PETase performs poorly. In contrast, a single W159A PETase mutant performed poorly in all temperatures, barely exhibiting any activity at 30°C.

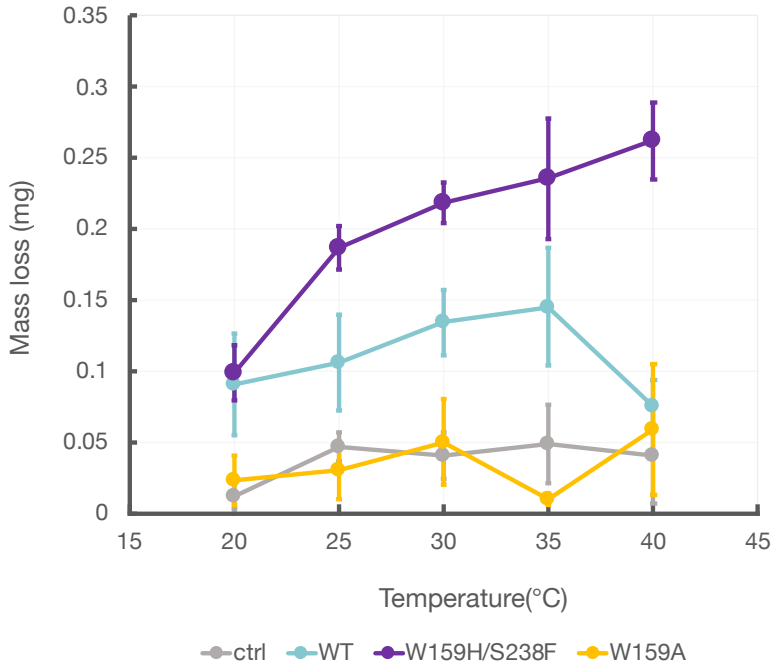


Figure 3-10: Effect of temperature on hydrolysis of PET after 4 days. Incubations took place in 50 mM sodium phosphate buffer (pH 7) at 30°C for 96 hours. Gray: control, Teal: WT PETase, Purple: W159H/S238F Yellow: W159A. Values shown are the mean of three replicates and error bars indicate standard error.

Thermophilic proteins have been shown to benefit from several factors – existence of disulfide bonding being one of the driving forces, for their thermostability[94]. On the other hand, non-covalent interactions, such as ionic interactions[95,96], salt bridging and hydrophobic interactions, shorter loops[97] and increased atomic packing [98] can also impact significant thermal stability; and there exist numerous hyperthermophilic proteins that have no disulfide bonds but withstand temperatures greater than 80°C[99,100]. It has been previously shown that PET degrading esterase function benefits from the addition disulfide bonds, for esterase enzymes exposed to high temperatures[101]. Furthermore, a PETase variant (S121E/D186H/R280A), which was reported to have a higher T_m value

(~8°C) than WT PETase exhibits tighter atomic packing within the β_5 β_6 β_7 strands of PETase. Taken together, we propose that the narrowed active site cleft of the PETase W159H/S238F double mutant may contribute to the enhanced thermal stability of this PETase variant, and thus leading to its ability to hydrolyze PET at higher temperatures than the optimum temperature of the WT PETase enzyme.

Surfactants

The effect of surfactants on cutinases have been documented previously[102,103]. While most surfactants interfere with cutinase enzyme activity, we have seen that the addition of SDS in very small concentrations enhances the activity of PETase and leads to increased mass loss of lcPET (when the latter is incubated with enzyme). At 0.02 w/v [SDS], mass loss of lcPET is almost 8-fold higher than 0% w/v [SDS]. At concentrations higher than 0.05 % w/v [SDS], there is however no significant mass loss, suggesting that the enzyme is inactivated by higher SDS concentrations (Figure 3-11 A). Meanwhile, addition of Triton X-100 in varying concentrations inhibits lcPET degradation by PETase, with no discernable mass loss even at lower Triton X-100 concentrations.

The staggering 8-fold increase in mass loss in the presence of 0.02 w/v [SDS] suggests that hydrophobic reactions between PET and SDS may be taking place, while SDS and PETase could be interacting via the negatively charged head groups of the SDS molecules. We additionally experimented with co-incubating PETase and MHETase enzymes with small concentrations of SDS, although no discernable mass loss in low concentrations of SDS versus no SDS was observed.

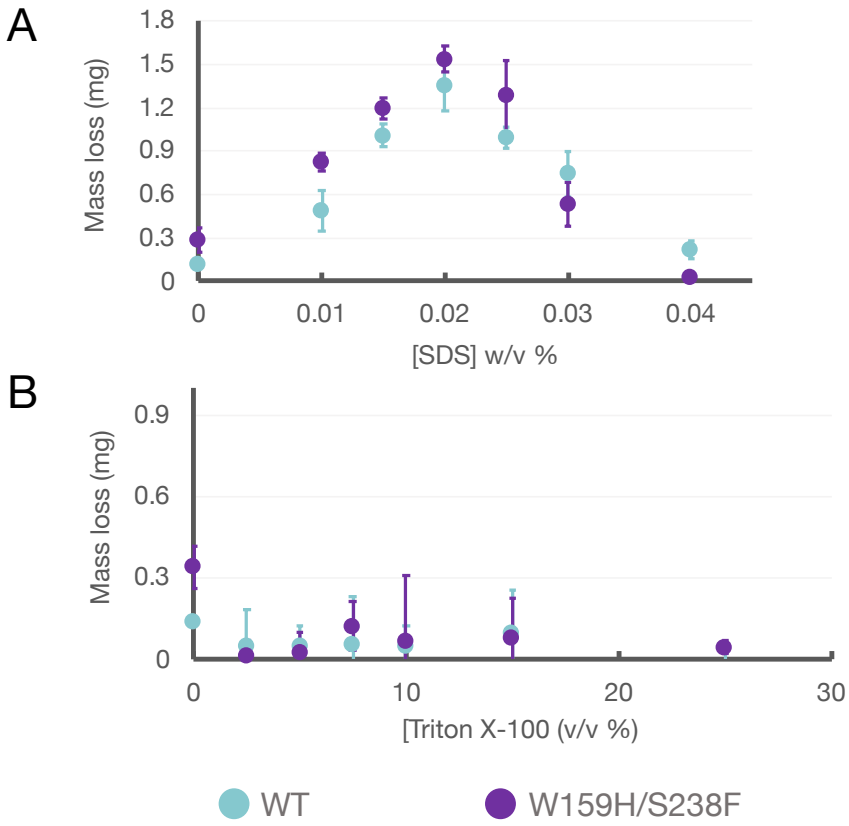


Figure 3-11: Effect of surfactants on lcPET mass loss. A) Addition of SDS and its effect on lcPET mass loss. B) Addition of Triton X-100 and its effect on lcPET mass loss. All incubations took place for 96 hours at 30°C in sodium phosphate buffer pH 7.0.

PETase Activity on Non-native Substrates

Hydrolysis of *p*-NP Esters

In relation to PET hydrolysis by cutinases, direct analysis methods such as HPLC provide quantitative information on PET degradation, although sample analyses can be a slow process. In order to more quickly assay esterase function, use of short chain and long chain esters of *para*-nitrophenol (*p*-NP) can be used. Its production rate at 405 nm can be easily monitored. As part of the esterase family of enzymes, PETase also possesses the ability to hydrolyze *p*-NP-esters of short to medium acyl chain lengths

including *p*-NP-acetate and *p*-NP butyrate (Figure 3-12, Table 3-3). The micromolar affinity of PETase for these substrates was found to be unexpected. Interestingly, WT PETase and W159H/S238F as well as W159A PETase single mutant performed similarly on *p*-NP acetate and *p*-NP butyrate substrates, with no significant difference in activity observed. When assessing the enzyme catalytic efficiency, i.e. K_m and k_{cat} values of all three proteins for the two esters, we also observed no significant difference between the different enzymes or substrates used for hydrolysis. However, argument could be made about how much the hydrolytic activity of PETase on *p*-NP acetate and *p*-NP butyrate substrates is representative of the action of PETase on an insoluble substrate, such as PET. This is especially true, since the action of the PETase W159H/S238F double mutant on PET has been shown to be significantly improved compared to WT PETase[2]. Nevertheless, using *p*-NP substrate for PETase enzyme assays provides a useful insight, as it can provide a readout of whether the proteins, expressed recombinantly, are folded correctly or not, and whether its catalytic site is active or not. Therefore, using *p*-NP esters as indicators of protein activity is desirable as an exploratory assay, but it cannot be an accurate indicator of PET hydrolysis efficiency.

Table 3-3: Kinetic measurements of WT PETase, W159H/S238F and W159A against *p*-NP-acetate and *p*-NP butyrate calculated using Graphpad Prism 8.0.

***p*-NP acetate**			***p*-NP butyrate**		
	Km (mM)	Kcat (s^{-1})		Km (mM)	Kcat (min^{-1})
PETase	~0.852	~411	**PETase**	~0.571	~421
W159H/S238F	~0.910	~275	**W159H/S238F**	~0.905	~302
W159A	~0.8494	~407	**W159A**	~0.8538	~415

Hydrolysis of BHET

A byproduct of PET hydrolysis by PETase, BHET is further hydrolyzed into MHET by PETase (Figure 3-12).18 hours post-incubation with PETase, BHET is depleted. Incubation of MHETase with BHET, however, yields no conversion to products whereas PETase and MHETase together completely convert BHET to TPA. We have further investigated what we refer to as the "BHETase" activity of WT PETase and the PETase W159H/S238F double mutant. WT PETase affinity for BHET was significantly higher compared to that of W159H/S238F (Table 3-4), shedding further light into how W159H/S238F may be slightly more effective at degrading PET than WT PETase, and suggesting that the double PETase mutant has a lower affinity for BHET, which would indicate that it is not as susceptible to a possible competitive inhibition (PETase has only one active site) by BHET.

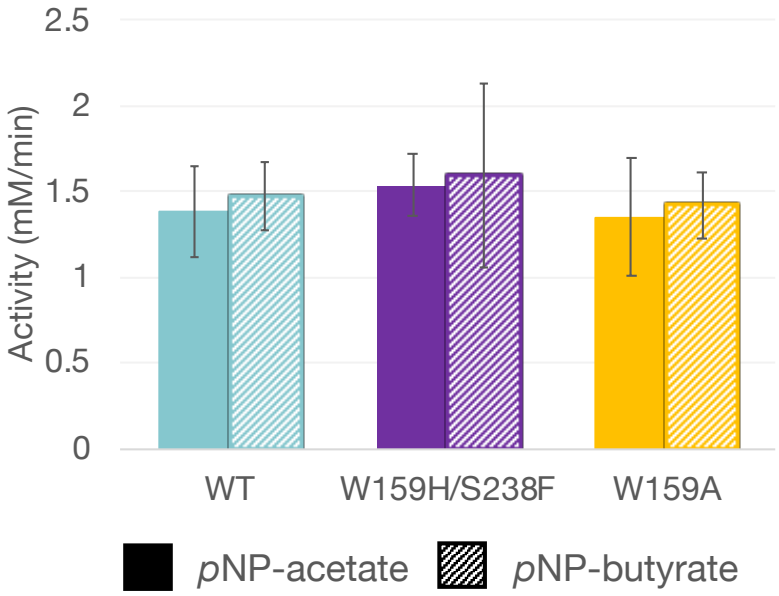


Figure 3-12: Hydrolysis of short chain *p*-NP esters by WT PETase (teal), W159H/S238F (purple) and W159A (yellow) calculated from linearity. Activities of enzymes against pNP-acetate are shown in solid fills, and activity against *p*-NP butyrate are shown as diagonal line fills. Activities are shown as the mean of three replicates and error bars indicate the standard error.

MHETase Activity on Non-native Substrates

While this section focuses on PETase activity, we additionally investigated the ability of the MHETase enzyme to degrade *p*-NP esters. MHETase, unfortunately, has no activity on *p*-NP acetate, *p*-NP butyrate nor *p*-NP octanoate. Initially, we suspected that MHETase might need to be interfacially activated to hydrolyze *p*-NP esters. The presence of DMSO in *p*-NP ester enzyme assays is necessary as *p*-NP acetate is unstable without at least 10% DMSO present. Furthermore, MHETase activity assays were already conducted in the presence of 10% DMSO (described in Chapter 2 and Chapter 4), an

organic solvent that is known to help with interfacial activation of lipases[104]. Addition of increased concentrations of DMSO in the buffer did not activate MHETase, ruling out the use of *p*-NP ester substrates for assaying MHETase. Additionally, MHETase appears to be very specific towards its substrate, MHET (Further described in Chapter 4), and we have not been able to identify a substrate that

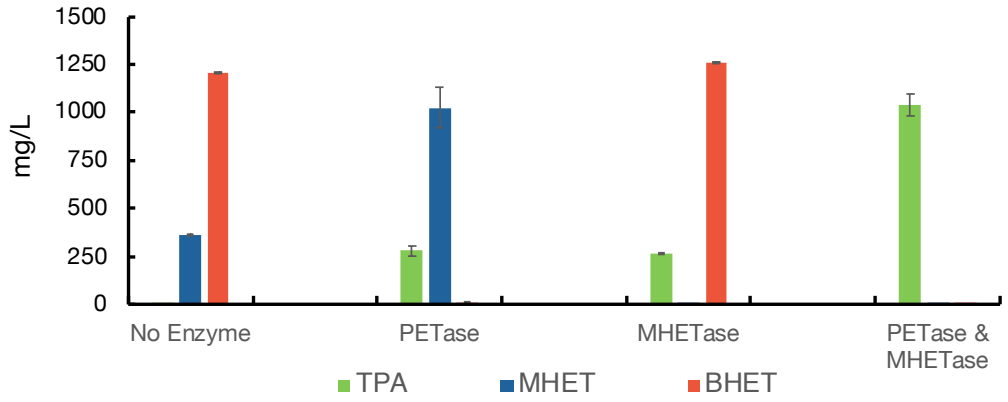


Figure 3-13: Initial hydrolysis experiment of BHET after 18 hours of incubation with no enzyme control, PETase only, PETase, MHETase and PETase and MHETase together. No enzyme control shows autohydrolysis of BHET. All reactions took place in sodium phosphate buffer pH 7.0, 10% DMSO and 1 g/L BHET. Values shown are the mean of three replicates and error bars show the standard error. Green: TPA, Blue: MHET, Red: BHET.

Table 3-4: Kinetic parameters of BHET hydrolysis. N/D: Not detected.

	K_{cat} (s^{-1})
PETase	0.9 ± 0.2
W159H/S238F	0.4 ± 0.1
MHETase	N/D

Exploratory Investigation of Putative PETases

Incubation of lcPET with several other putative PETases such as *Rhizobacter gummiphilus* putative PETase, *Pseudomonas saudimassiliensis* putative PETase, and *Pseudomonas bauzanensis* putative PETase has unfortunately yielded no lcPET degradation products. Taking these results into consideration, and the fact that PETase from *I.* sakaiensis is an exclusively aromatic polyesterase, we investigated whether these aforementioned putative "PETases" can degrade PBS and/or PLA (Table 1-1); yet no degradation products were observed via SEM resulting from the enzyme assays with these two substrates.

Concluding remarks

Following the discovery of *I.* sakaiensis by Yoshida and colleagues, enzyme engineering efforts focused on increased hydrolysis of PET by PETase gained significant attention from the public and researchers alike. High quality structural and experimental information has been generated about PETase (and many PET degrading cutinases (Table 1-1)) as a result of these efforts. It is also widely recognized in the research field of enzyme-catalyzed solid substrate degradation that chemistry alone cannot completely account for the intermolecular and nano-scale events taking place during these processes[105]. SEM analysis of PETase on lcPET coupons revealed preferential hydrolysis patterns associated with the order and crystallinity of the substrates. Furthermore, the pattern of PETase hydrolysis on sPET revealed the presence of spherulitic structures on

the substrate surface, accompanied by a decrease in crystallinity post-treatment of sPET substrate with PETase. To further explore the crystallinity change phenomena in hydrolyzed substrates, we stained ultrathin sections of lcPET and boPET with RuO_4 and examined their lamellar width pre-and post-incubation. This novel method for visualizing PET hydrolysis revealed an increase in percent order in lcPET, whereas boPET displayed a slight decrease in polymer chain order. While high crystallinity PET exhibited change in morphological features, further assessment of change in chain order needs to be conducted with not only boPET, but also sPET.

The W159H/S238F PETase double mutant displayed improved performance over WT PETase at higher temperatures (i.e. 40ºC). The PETase W159H/S238F double mutant protein was designed with the idea that it would be comprised of a narrower active cleft than WT PETase, similar to that of *Tf*Cut. In this case, increased atomic packing might help explain the increased thermal stability of W159H/S238F. Moreover, engineering PETase to be more like *Tf*Cut, a cutinase from a thermophilic organism, might explain why the PETase W159H/S238F double mutant is more thermally stable. Furthermore, 40°C is closer to the T_g of PET. This might also be facilitating the enzyme's access to the PET polymer chains.

Assaying the activity of WT PETase, W159H/S238F and W159A against *p*-NP acetate and *p*-NP butyrate revealed an interesting feature and that their activities on both of these esters are very similar. However, their activities on solid substrate PET is different, indicating that *p*-NP esters, while can help with assessing whether the active

site of the protein is folded properly or not, are not a reliable indicator of PET hydrolysis capability.

PETase was also shown to hydrolyze BHET to some extent. Kinetic measurements of WT PETase and W159H/S238F showed that the PETase W159H/S238F double mutant has a lower affinity for BHET than the WT enzyme and suggests that the double mutant may be less susceptible to competitive inhibition by BHET. Meanwhile, we detected no MHETase activity when the enzyme was incubated with BHET, which was unexpected. The structural and chemical basis for this peculiar behavior is described further in Chapter Four, focused solely on MHETase.

CHAPTER FOUR

STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF MHETASE

Contribution of Authors and Co-Authors

Author: Brandon C. Knott*

Contributions: Computational work and molecular simulations

Co-Author: Ece Topuzlu *

Contributions: Protein purification, assays, microscopy and synergy studies

Author: Mark D. Allen *

Contributions: Crystallographic work

*These authors contributed equally to this work.

Co-Author: Japheth Gado

Contributions: Bioinformatics work

Co-Author: Harry P. Austin

Contributions: Crystallographic work

Co-Author: Erika Erickson

Contributions: Substrate assays and HPLC work

Co-Author: Bryon S. Donohoe

Contributions: Microscopy and substrate assays

Co-Author: Nicholas A. Rorrer

Contributions: Synthesis of substrates

Co-Author: Fiona L. Kearns

Contributions: Molecular simulations

Co-Author: Graham Dominick

Contributions: Cloning and molecular biology

Co-Author: Christopher W. Johnson

Contributions: Cloning and molecular biology

Co-Author: Valérie Copié

Contributions: Experimental design, manuscript edits

Co-Author: Christina M. Payne

Contributions: Bioinformatics design

Co-Author: H. Lee Woodcock

Contributions: Wrote the first draft of the manuscript

Co-Author: Gregg T. Beckham

Contributions: Wrote the first draft of the manuscript

Co-Author: John E. McGeehan

Contributions: Wrote the first draft of the manuscript

Manuscript Information Page

Structural and biochemical characterization of MHETase

Brandon C. Knott*, Ece Topuzlu*, Mark D. Allen*, Japheth Gado, Harry P. Austin, Erika Erickson, Bryon S. Donohoe, Nicholas A. Rorrer, Fiona L. Kearns, Graham Dominick, Christopher W. Johnson, Valérie Copié, Christina M. Payne, H. Lee Woodcock, Gregg T. Beckham, John E. McGeehan

*These authors contributed equally to this work.

Proceedings of the National Academy of Sciences of the United States of America

Status of Manuscript:

__X_Prepared for submission to a peer-reviewed journal
____Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

National Academy of Sciences

Abstract

The accumulation of synthetic plastics in the biosphere is leading to a global environmental crisis. In response, microbes are evolving strategies to convert man-made polymers into carbon and energy sources, and these systems offer a promising starting point to harness for biotechnological purposes towards plastics upcycling. To that end, *Ideonella sakaiensis* 201-F6 was recently reported to secrete a two-enzyme system to deconstruct the abundant synthetic polyester, polyethylene terephthalate (PET), to its constituent monomers for further catabolism. The PETase enzyme specifically breaks down the PET polymer, liberating mono-(2-hydroxyethyl) terephthalate (MHET), which is then cleaved to its constituent monomers, terephthalic acid and ethylene glycol, by MHETase. Here we report the MHETase crystal structure solved at 1.6 Å resolution, which illustrates that the core domain of MHETase is highly similar to PETase, but with a "lid", and that overall, the MHETase sequence and structure closely resembles tannase enzymes. We have used molecular dynamic simulations to show that MHETase follows the canonical serine hydrolase mechanism. In addition, we have produced multiple mutants of MHETase and PETase to demonstrate that the lid domain of MHETase is important for function, and that homologous tannases from two other organisms also turnover MHET. Interestingly, the enzyme does not turnover mono-(2-hydroxyethyl) isophthalate (MHEI), but docking simulations suggest potential amino acid mutations that would lead to turnover of this compound, which is also present in most industrial PET samples. Lastly, we have identified an optimal ratio of PETase and MHETase to

maximize synergy between the two enzymes in PET digestion experiments, which is a mass ratio of PETase:MHETase of 1:2. Taken together, these results offer new molecular-level insights into the two-enzyme system of *I. sakaiensis* in the breakdown of recalcitrant PET.

Significance

Plastics have become essential to modern lifestyles, but due to their widespread abundance, low cost, and durability, it is now widely accepted that they are creating a worldwide pollution crisis. In response to the rise of plastic pollutions in soils, air, and water, microbes are evolving the ability to consume synthetic plastics. Understanding how life is evolving in response to plastic exposure will accelerate the development of robust industrial and biotechnological solutions to depolymerize and upcycle plastics, an approach towards sustainability, ecosystem protections, and the concept of a global circular economy. In this work, we elucidated new details of a how a two-enzyme system deconstructs polyethylene terephthalate, the most abundantly produced polyester, which is ubiquitously used in single-use beverage bottles, carpets, and clothing, and other synthetic fiber products.

Introduction

Synthetic polymers permeate all aspects of modern society, due to their low cost, high durability, and impressive extents of tunability. Originally developed to avoid the

use of animal-based products, plastics have now become so widespread that their leakage into the biosphere and accumulation in landfills is creating global-scale environmental crises. Indeed, plastics have been found widespread in the world's oceans[106–109]in soil[110], and more recently, microplastics have been reported even to pollute the air and clouds [111]. The leakage of plastics into the environment on a global scale has led to the subsequent discovery of multiple biological systems that are able to utilize man-made polymers for use as carbon and energy sources [1,112–116]. On the basis of these natural systems being able to degrade synthetic plastics, the environmental microbiology community has begun examining how natural enzymes evolve to consume non-natural substrates, which in turn could enable the use of these enzymes and their plastic degradation capabilities to be used for biotechnological applications towards a more circular materials economy , wherein the historical narrative of disposing and discarding of postconsumer materials is replaced by regeneration and "upcycling" of postconsumer waste[116–120]

Among synthetic polymers manufactured today, polyethylene terephthalate (PET) is the most abundant polyester, which is made from petroleum-derived terephthalic acid (TPA) and ethylene glycol (EG). Likely due to the prevalence of esterase enzymes present in nature, the biodegradation of PET has long been studied, with multiple cutinase enzymes reported to catalyze its breakdown [62,121–127]. In 2016, Yoshida et al. reported the discovery and characterization of the soil bacterium, *Ideonella saikaiensis* 201-F6, which employs a two-enzyme system to depolymerize PET to TPA and EG, which are further catabolized to be used as major carbon and energy sources [1].The detailed characterization of this bacterium revealed that the PETase enzyme, which is a

cutinase-like serine hydrolase that attacks the PET polymer, liberates mono-(2-hydroxyethyl) terephthalate (MHET). This soluble product is further hydrolyzed by the action of a second enzyme called MHETase to generate TPA and EG. Multiple crystal structures and protein engineering studies have been reported to date for the *I. sakaiensis* PETase enzyme [2,66–68,128–130]and have revealed an open active site protein architecture that is able to bind to PET oligomers. The PETase enzyme very likely follows the canonical serine hydrolase catalytic mechanism [131], but open questions remain regarding the mobility of certain residues and the functional significance of amino acid flexibility during the catalytic cycle [67].

Conversely, the MHETase enzyme has been far less studied and less understood to date. Palm et al. published an elegant structural and engineering study, which represents the only reported crystal structure of MHETase from *I. sakaiensis* to date [76]. Therein, the authors reported a 2.2 Å resolution structure, whose scaffold resembles that of ferulic acid esterase enzymes (FAE) [132,133]. Using this novel structure as a guide for engineering site-directed protein variants, the authors successfully generated a mutant of the wild type enzyme which can hydrolyze bis-(2-hydroxyethyl) terephthalate (BHET) by MHETase, for which BHET is not a natural substrate. Despite this published report and the initial report of MHETase from Yoshida et al. [1], multiple questions remain regarding the mechanism of MHETase action and PETase-MHETase enzyme synergy in PET degradation.

To that end, we report in this study several new crystal structures of ligand-bound MHETase enzymes, including a 1.6 Å resolution structure of MHETase containing a

bound benzoate ligand. The structure reveals that MHETase and PETase share a common α/β hydrolase core domain, with MHETase exhibiting a "lid"-like structure and a semi-buried active site, similar to FAE [133]. The high-resolution structure with a benzoate ligand bound also enables the computational study of a putative two-step acylation-deacylation reaction mechanism, using mixed quantum mechanics/molecular mechanics (QM/MM) approaches. These computer simulations indicate that the deacylation is the rate-limiting step, and that the EG leaving group rapidly exits the catalytic pocket, following product formation. Detailed bioinformatics analyses also highlight two similar enzymes to MHETase, which we experimentally confirmed as being active on the MHET substrate. Multiple mutants of MHETase were also engineered to explore the effect of the presence of disulfide bonds on enzyme activity; the effect of the lid on both PETase and MHETase; and the role of several highly conserved residues on the catalytic mechanism of MHETase. Additionally, MHETase activity on mono-(2-hydroxyethyl) isophthalate (MHEI) and mono-(2-hydroxyethyl) furanoate (MHEF) subtrates were also tested. Lastly, PETase-MHETase synergy experiments were conducted using PET as substrate, and which revealed that an approximately 1:1 molar ratio of PETase:MHETase is optimal for PET degradation. Overall, this study provides new molecular insights into MHETase enzyme structure and function and reinforces the concept that PETase and MHETase are highly related enzymes with similar protein core structures and catalytic mechanisms, but tuned to bind to solid and soluble substrates, respectively.

Results

Structural Characterization of MHETase

As a serine hydrolase, the catalytic domain of MHETase adopts a typical α/β-hydrolase fold (Figure 4-1A, light gray), in addition to containing a lid domain (Figure 4-1A, dark gray) that partially covers the enzyme active site residues. Herein, we present the highest resolution structure of MHETase to-date, solved to a 1.6 Å resolution and containing a benzoate molecule bound at the active site.

Like its companion PETase, MHETase exhibits a conserved Asp-His-Ser catalytic triad (Asp492, His528, Ser225) (Figure 4-1B). The catalytic residues of MHETase align structurally with those of PETase, including the loops containing the catalytic triad residues in the immediate vicinity of the active site. A good structural alignment is observed in several regions overall. A large helix (~15 residues) beginning at MHETase N195, and corresponding to PETase S125, whose positions align closely between the two enzymes. After some structural sequence divergence, parallel β strands extend from MHETase K218 (PETase A152) through MHETase C224 (PETase G158) leading to the catalytic serine (MHETase S225; PETase S160), which begins a relatively 14-residue long α helix. Structural alignment persists until about MHETase residue Y252 (PETase D186). PETase residues 189-238 align closely with MHETase residues 470 to 529, which includes a catalytic histidine as well as a histidine-coordinating catalytic aspartate.

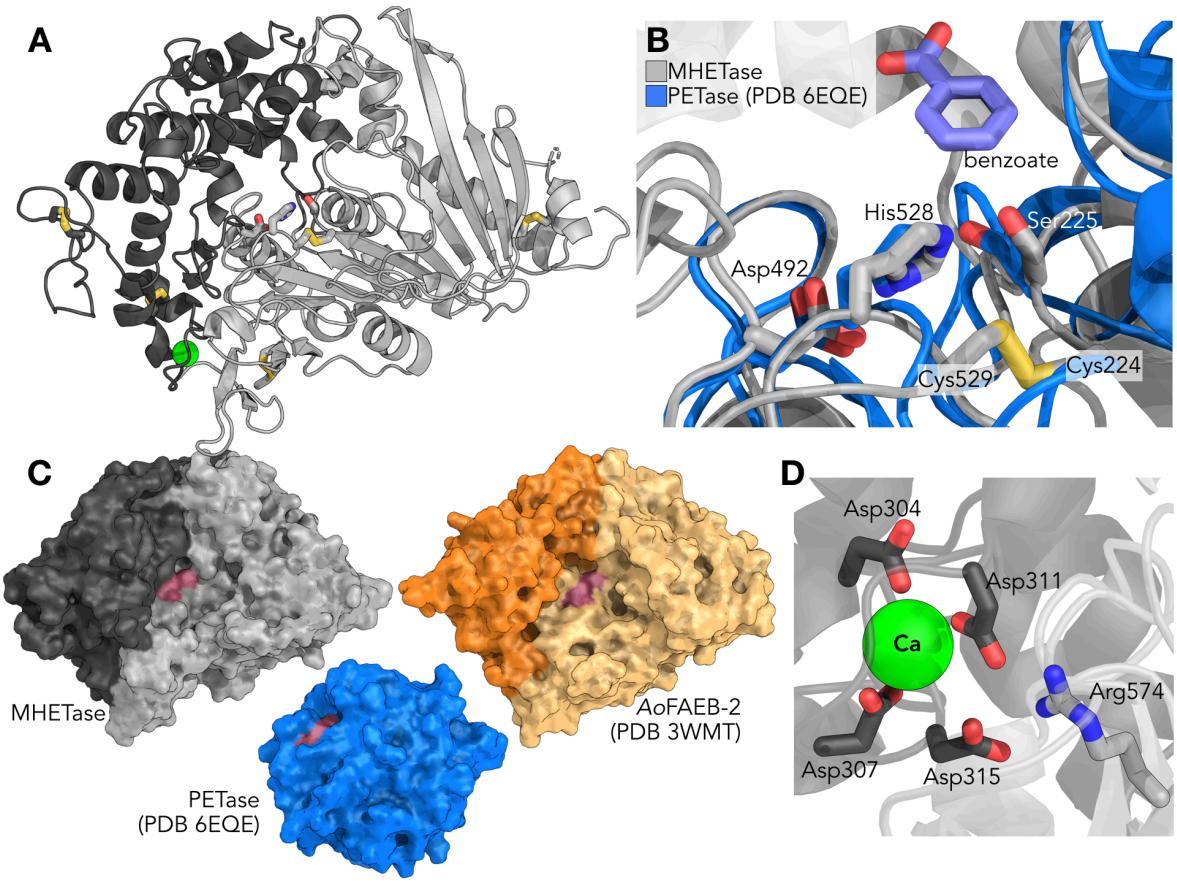


Figure 4-1: Structural analysis of MHETase. **A**) overall MHETase structure (resolution 1.6 Å) showing catalytic triad, five disulfides, and calcium ion. The lid domain is shown in dark gray, whereas the catalytic (non-lid) domain is shown in light gray. **B)** Catalytic triad of MHETase (gray sticks) shown with bound benzoate (slate sticks) demonstrating the alignment of the catalytic residues with those of PETase (blue sticks). In addition, the active site disulfide of MHETase is shown in gray sticks. **C)** Lid domains of MHETase and Aspergillus oryzae FAEB-2 are shown in dark gray and dark orange, respectively. IsPETase, which lacks this domain, is shown in blue surface. The catalytic triad is shown in all cases in raspberry surface**. D)** MHETase calcium binding site.

In the largest structural divergence from PETase, MHETase is comprised of a "lid" domain, a structural motif found in tannase family of enzymes (Figure 4-1C). In MHETase, the lid domain spans a large stretch of amino acid residues, starting from Pro250 through Thr472. Intriguingly, the connection points for the lid domain align reasonably well with PETase Pro184 and Ser192, but in PETase, these are connected by a

small seven-residue loop. In this respect, the overall structure of MHETase is most similar to ferulic acid esterases (FAE). In *Aspergillus oryzae* FaeB (AoFaeB, PDB ID: 6G21), this lid domain covers the active site domain, as in MHETase[133]. Much like AoFaeB, the lid domain helps form the substrate binding pocket.

MHETase contains five disulfide bonds, while FAE enzymes generally contain six disulfide bonds. In AoFAEB-2, one Cys of the 6th disulfide is a single residue variation with MHETase, whereas the other Cys is located on a loop that constitutes a ~15 residue stretch which is absent in MHETase. One of the MHETase disulfide bonds is located within the active site, connecting cysteines (Cys224 and Cys529) that are adjacent to the catalytic residues Ser225 and His528. This disulfide bond is completely conserved in tannase family members. PETase lacks this disulfide bond, but in a fascinating twist, the two equivalent residues are the same two residues that were mutated in PETase by Austin et al. [2]to create a substrate binding groove which is more similar to cutinases, and which resulted in improved enzyme catalytic activity on crystalline PET.

Contained within the lid domain of MHETase is a calcium binding site. The calcium ion is coordinated by four aspartic acid residues, consisting of Asp304, Asp307, Asp311, and Asp315. Fourth aspartate, Asp315, in MHETase is replaced by a glutamate (E) residue in AoFaeB. Asp304, Asp307, Asp311 are highly conserved residues among the 6,000 sequences in the tannase family. Asp315 in MHETase is much more variable and is only the third most common residue seen in the rest of the tannase family members. On the other hand, calcium binding site-adjacent residue Asp574 is extremely conserved in all these sequences. (Figure 4-1D).

Bioinformatics Analysis of MHETase Sequences and Structural Homology

Ideonella sakaiensis MHETase (*Is*MHETase)belongs to the tannase family of enzymes (PFAM ID: PF07519) which consists of fungal and bacterial ferulic acid esterases, fungal and bacterial tannases, as well as several bacterial homologues of unknown function [134]. Based on substrate specificity and phylogenetic relationships, the very diverse group of ferulic acid esterases have been classified into 13 sub-groups, which includes tannase family ferulic acid esterases (group SF-11) and other nonhomologous groups [135]. Similarly, tannases of the tannase fungal family are nonhomologous to the tannases which belong to tannase_bact family. Our analysis of 6,671 tannase family sequences shows that *Is*MHETase shares low sequence similarity with most sequences in the the tannase enzyme family members ((Figure 4-7), with the exception of two close homologs of *Comamonas thiooxydans* and *Hydrogenaphaga* sp., which share identities with *Is*MHETase of 81% and 73%, respectively. All other entries in the tannase family share sequence identities of less than 53%.

We have conducted a phylogenetic analysis of 120 selected tannase family sequences, including *Is*MHETase, to illustrate the functional and taxonomical diversity in the tannase family. In the phylogenetic tree (Figure 4-8), bacterial and fungal enzymes form separate groups, and within these groups, ferulic acid esterases (FAEs) and tannases form separate sub-groups. This indicates that tannases and ferulic acid esterases within the broad tannase family of enzyme have evolved from a common ancestor. *Is*MHETase, along with the *Comamonas* and *Hydrogenaphaga* homologs, forms a sub-clade within a group of proteobacterial ferulic acid esterases (bootstrap values > 95%). This suggests

that MHETase has evolved from ferulic acid esterases, rather than tannases. Further supporting this hypothesis is the fact that *Is*MHETase has a higher alignment score when aligned with the profile hidden Markov model (profile HMM) of ferulic acid esterases (456.8) compared to when aligned with the tannase profile-HMM (396.8). Taken together, phylogenetic analysis and the low similarity of *IsMHETase* to other tannase family sequences suggest that the emergence of MHET-hydrolase activity is relatively recent event in the evolution of tannase family enzymes.

Using a multiple sequence alignment of 6,671 tannase family sequences, we have carried out a residue position conservation analysis with residue positions in *Is*MHETase as reference point (Figure 4-9A). As expected, most positions in the active site are strongly conserved. However, among the residues located within the active site, Ala257, Arg411, Phe415 and Ser416 of *Is*MHETase have the lowest conservation scores when analyzed against the tannase family (Figure 4-9 B&C). The variability of these four residue positions suggests that they may be promising targets for mutagenesis to expand the specificity and activity of MHETase. Indeed, in a recent work[76], significant improvements of *Is*MHETase activity on MHET and BHET substrates were observed, when residues at by positions 411, 415, and 416 were mutated. Further probing and engineering of these positions in *Is*MHETase may yield improved result regarding enhanced catalytic performance of the enzyme. Furthermore, the 10 Cys positions in *Is*MHETase which form 5 disulfide bonds are all strongly conserved in the tannase enzyme family (Figure 4-11) Although a 6th disulfide bond exists in the *Aspergillus*

oryzae ferulic acid esterase (AoFAEB-1), less than 8% of tannase family sequences utilize a 6th disulfide, and the 6th disulfide positions are extremely variable (Fig 4-11).

Biochemical Characterization and Engineering.

Toward the goals of tuning enzyme activity, enhancing substrate specificity, and thermostability of enzymes for plastics degradation, we produced a suite of engineered enzymes mainly through engineering of the lid and various disulfide bonds. A full list of protein mutants that have been engineered can be found in Table 4-1, Table 4-2, and Table 4-3.

In addition, we purified putative MHET-degrading enzymes from *Hydrogenaphaga* sp. and *Comomonas thioxydans*, which were identified via an exhaustive search of tannase family member sequences.

Molecular simulations and docking

To further understand the dynamics, substrate binding, and catalytic mechanism of MHETase, molecular dynamics simulations of MHETase were conducted, including a nearly one microsecond calculation using classical molecular dynamics (MD). These calculations included three independent simulations each of 100 ns duration for the apo enzyme, enzyme with Ca^{2+} bound, and enzyme with both Ca^{2+} and MHET subtrate bound. The active site configuration with bound MHET is shown in Figure 4-2A.

When MHET is bound at the active site, its carboxylate motif is tightly bound by interactions with the sidechains of residues R411 and S416, forming three persistent hydrogen bonds throughout 300 ns of MD simulation (Figure 4-2). Stable hydrogen

bonds are also formed between the oxyanion hole formed by Glu226 and, to a far greater extent, by Gly132. These constitute the primary hydrogen bonding interactions of bound MHET in the active site of MHETase.

Gly132 resides on a loop (Gly129-Ala150) that, in the apo enzyme, is one of the most flexible regions of the enzymes (Figure 4-2B, outlined in blue). With MHET bound, however, the loop is significantly stabilized. This loop also includes Ser136; a point substitution at this position to a cysteine residue constitutes half of the relatively unique sixth disulfide bond that is present in AoFaeB. Replacing the MHETase loop (W75-H91) with a longer loop from AoFaeB (Y60-F89 in that enzyme, which includes C76) would introduce this sixth disulfide into MHETase, thus linking two regions that appear to be highly flexible in MHETase. If this sixth disulfide were introduced in analogy with AoFaeB, we expect that this would most certainly significantly stabilize these two specivally mobile regions of MHETase and may potentially increase the thermostability of the enzyme.

In AoFAE-B, the calcium ion bound in the lid domain far from the active site has been proposed to have a stabilizing role on the lid domain [133]. Our simulations of MHETase with and without bound calcium ion provide a test of this hypothesis. RMSF results comparing the fluctuations of backbone atoms show significant differences mainly in the region directly adjacent to the calcium binding site (residues 304 to 315, Figure 4-2 B,

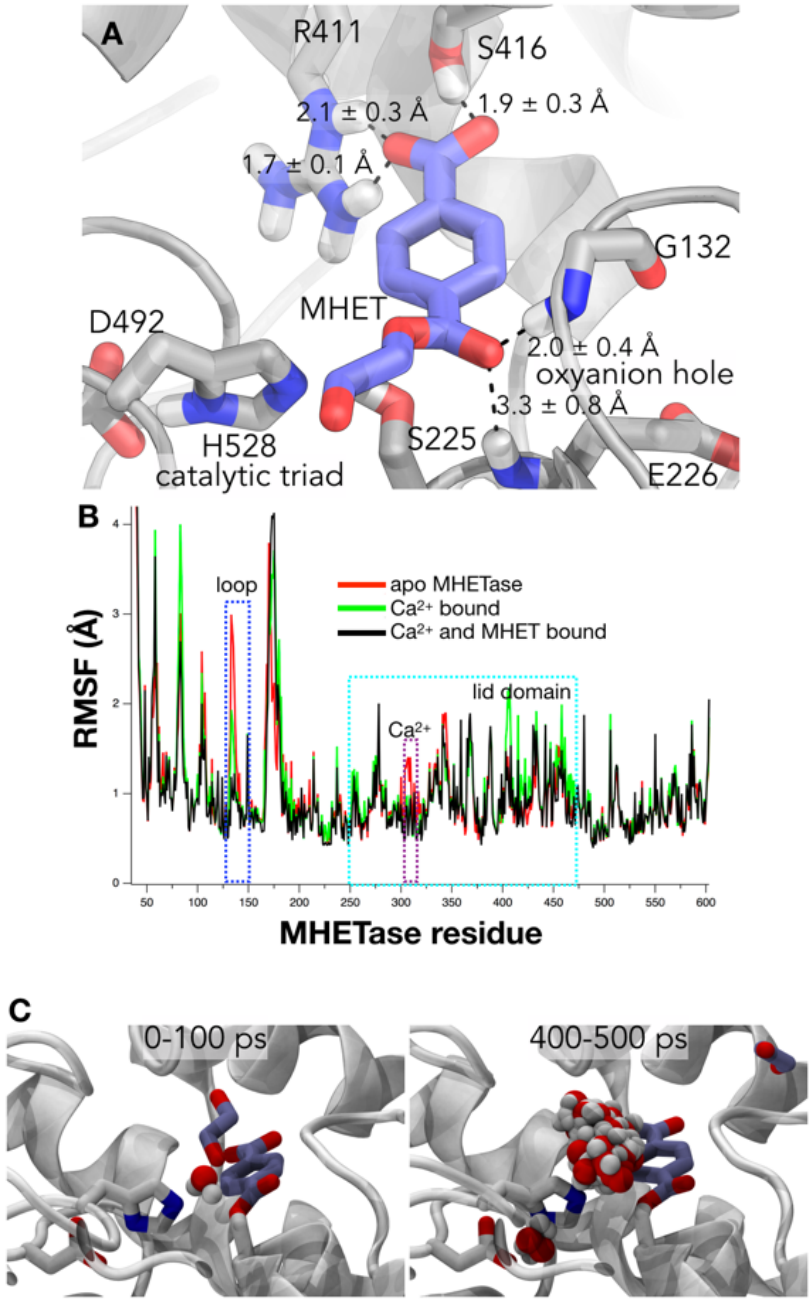


Figure 4-2: A) Active site configuration of MD simulations and hydrogen bonding distances of active site residues with MHET. B) Root mean square fluctuations (RMSF) of MHETase residues (side chain and backbone atoms). C) Number of waters within 2.5 Å of the charged nitrogen of catalytic histidine His528, visualized every 2 ps in the first 100 ps following acylation (wherein EG is still present in the active site) and 300 ps later (after EG has left the active site).

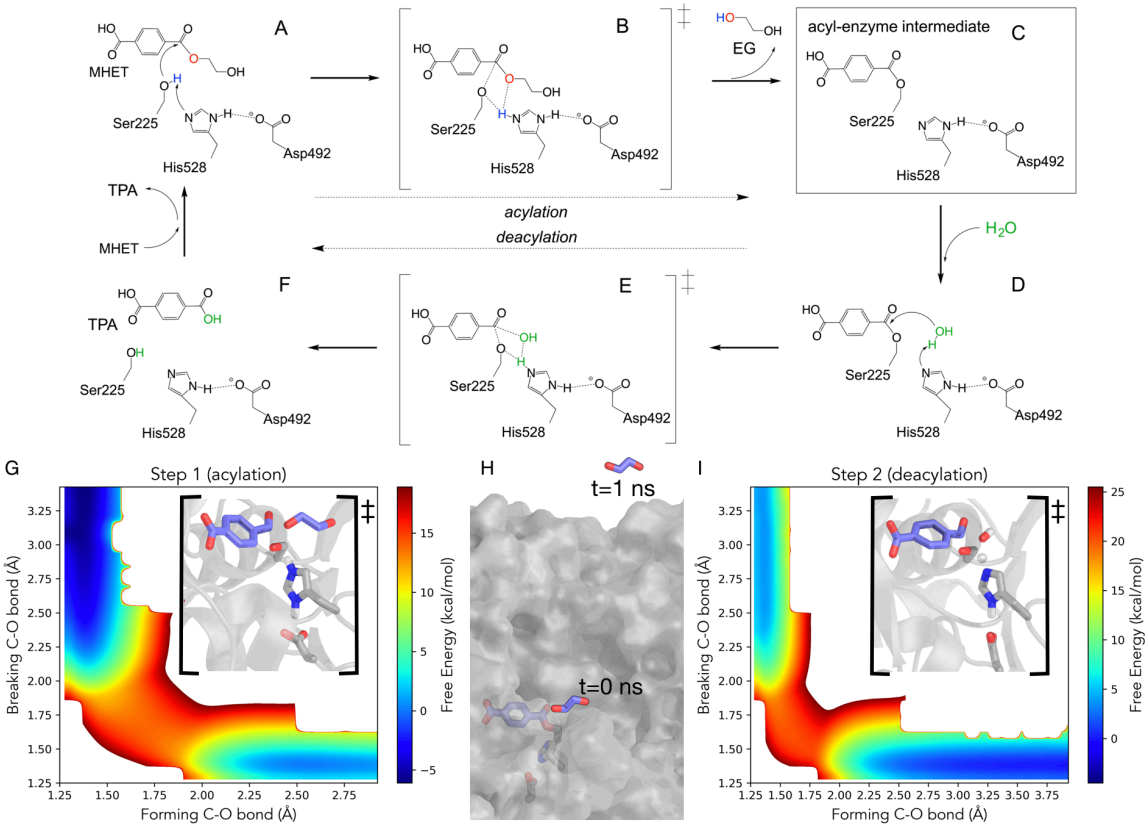


Figure 4-3: The catalytic mechanism of MHETase. A) reactant, B) transition state, and C) product of acylation in which His528 transfers a proton from Ser225 to the ethylene glycol (EG) leaving group. In the second step (deacylation, panels D- F), His528 plays a similar role and restores the catalytic serine by transferring a proton from a water molecule to Ser225 and generating a free terephthalic acid (TPA). G) Free energy surface for acylation computed along the breaking and forming carbon-oxygen bonds. H) Following acylation, EG leaves the active site, demonstrated via MD simulation. I) Free energy surface for diacylation reaction, demonstrated to be the rate-limiting step of the catalytic mechanism.

in purple); other regions of the lid appears to be unaffected by calcium binding, and the molecular dynamic simulations of the enzyme with calcium bound display even larger fluctuations in many regions of the lid domain (residues 250 to 472, Figure 4-2B, highlighted in cyan), suggesting the calcium does not significantly influence the stability of the lid domain.

In addition to classical MD simulations, hybrid quantum mechanics / molecular mechanics (QM/MM) simulations were performed to probe the acylation (Figure MD2A-C) and deacylation (Figure 4-3) catalytic steps. The acylation reaction reveals a 13.9 kcal/mol free energy barrier and a ΔG of reaction of -6.1 kcal/mol (Figure 4-3G). The deacylation reaction exhibits a free energy barrier of 19.5 kcal/mol and a ΔG of reaction of 3.4 kcal/mol (Figure 4-3I). Taken together, the two catalytic steps are overall favorable, resulting in a negative free energy of -2.7 kcal/mol. The rate-limiting step is indicated to be the diacylation step, with a rate of 0.12 s-1 (assuming that transition state theory is valid and a transmission coefficient of unity), nearly four orders of magnitude slower than acylation. These results indicate that MHETase follows a canonical serine hydrolase mechanism of substrate conversion to products[131].

Following the cleavage of EG from MHET during acylation, MD simulation reveals that the EG product leaves the active site within 1 ns (Figure 4-3 H). Immediately following cleavage, EG maintains a hydrogen bond with the catalytic histidine (H528) for approximately 100 ps, then dislodges from the active site. As it moves away from the enzyme active site, EG is fleetingly caught by interactions with a trio of serine side chains (S131, S136, and S138) but is completely free in solution within 800 ps. One significant implication of this process is that the deacylation reaction proceeds without EG being bound in the active site. This allows significantly more water molecules to approach the charged nitrogen of the catalytic histidine, offering enhanced opportunity for catalytic activation of the rate-limiting step. (Figure 4-2C).

MHETase synergistic interaction with PETase

PETase activity is inhibited by its products, BHET and MHET. As a result, we investigated whether coincubation of PETase with MHETase could improve PET turnover. At an optimal ratio of PETase:MHETase mass loading of approximately 2:1 (corresponding to a ~1:1 molar loading), PET mass loss was doubled over using PETase only. At loadings of 0.6 and 0.3 mg/g PET of MHETase and PETase, respectively, a mass loss of 0.31 mg was observed from a 10-13 mg PET coupon.

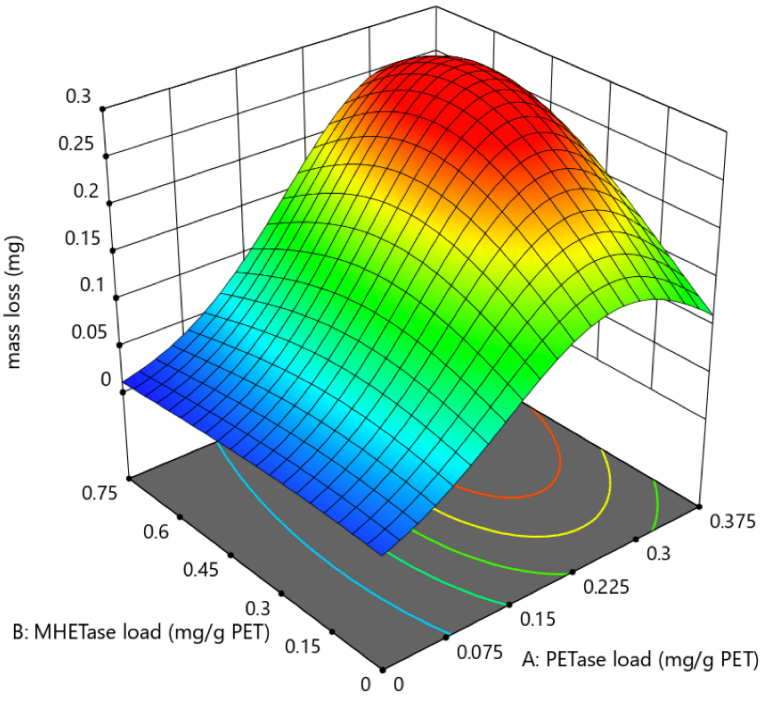


Figure 4-4: Response surface of the synergistic action of PETase and MHETase on lcPET. The two independent factors, PETase loading (mg/g PET, axis A) and MHETase loading (mg/g PET, axis B) are modeled as a function of mass loss of lcPET coupons. All lcPET and enzyme(s) were incubated in 50mM sodium phosphate buffer at pH 7 in 30°C for 96 hours.

MHETase activity on non-native substrates

Toward the goal of evaluating the substrate specificity of MHETase, activity assays were performed with mono (hydroxy 2-ethyl) isophthalate (MHEI). Over the course of 24 h, no detectable activity is detected, with substrate concentrations ranging from 25-250 μM MHEI; this is contrasted with nearly complete turnover of MHET with identical time period and initial substrate concentration (Table 4-8). Our attempts of incubation of MHETase with MHEI, however, did not yield any product.

Discussion

Whereas PETase works at the liquid-solid interface, MHETase works on a soluble degradation product of PET. The enzyme requirements for solid vs. soluble substrates are reflected in their architecture: PETase has a relatively flat binding surface and MHETase present a curved / "cup" structure, which is likely to be more effective at capturing free floating MHET molecules in solution.

The lid structure of MHETase also explains why it is not able to degrade PET. The six known MHET-hydrolytic enzymes that were reported by Yoshida et al. also hydrolyze PET [1]It is possible that MHETase cannot degrade PET because its structure presents steric hindrance which would prevent the binding of the PET polymer in the active site of MHETase.

Palm et al. noted the importance of the charged interactions between R411 and the carboxylate motif of the bound ligand [76]. Yet their structure MHETase bound to ligand

did not reveal the presence of hydrogen bonding between Ser416 and the substrate, while our crystal structures of MHETase with a benzoate ligand does indicate that hydrogen bonding between S146 and substrate is indeed present and important.

Palm et al. demonstrated the mutation of Arg411 abolished MHET activity, but conferred BHET activity. This change in substrate specificity was attributed to a change in the chemical nature of the substrate near Arg411 (from charged carboxylate to neutral alcohol motif). Steric impacts may be at play as well. Binding of BHET via interaction with R411 of the enzyme likely pushes the ester motif out of the active site past the catalytic triad, whereas shorter side chains may allow for BHET binding without the need to create hydrogen bonding interactions with Arg411. MHETase crystal structures with and without MHETA substrate analogues revealed a significant conformational change in F415 over the active site. This effect is not seen in MD simulations, despite similar binding mode of MHET (in simulation) with the MHETA (crystal structure). In fact, simulations show F415 to be more flexible when substrate is not bound. This difference in MHETase behavior could be attributed to the difference in ligand (MHET / MHETA) or time scale (the effect may occur on a timescale longer than O (100 ns)).

Within the alpha/beta hydrolase family of serine hydrolases, either the acylation or the deacylation step may be rate-limiting [131]. Here, we show that the second step of deacylation is the rate-limiting step by nearly four orders of magnitude. The deacylation reaction rate as calculated from QM/MM free energy barrier and TST is lower than experimental measurements by about two orders of magnitude (30 s^{-1} [1,76]), QM/MM

studies of the catalytic reaction steps of MHETase reveal that both steps proceed in a concerted/dynamic fashion.

One of the contributing factors to efficient PET hydrolysis is product inhibition by BHET and MHET [136]Alleviation of product inhibition by coincubation of PETase and MHETase reveals that at a molar ratio of 1:1, mass loss of lcPET can be doubled, compared to PETase alone. We have found no protein-protein interactions between PETase and MHETase, but it appears that the presence of MHETase positively impacts PET hydrolysis.

Conclusions

While PETase has gained significant attention from the research community, MHETase activity, substrate specificity and mechanism has stayed in the shadows. Our work sheds light on the far less studied counterpart of PET hydrolysis, MHETase, and reveals potential protein engineering aspects that can be further exploited for a more efficient method of polyester recycling.

Author contributions

GTB and JEM designed the study; BCK, ET, MDA, JG, HPA, EE, BSD, NAR, FLK, GD, CWJ conducted the experimental and computational work; all authors analyzed the data; BCK, HLW, GTB, and JEM wrote the first draft of the manuscript, which was edited and approved by all authors.

Acknowledgements

GTB, BCK, EE, BSD, NAR, GD, and CWJ thank the NREL Laboratory Directed Research and Development program for funding. JM thanks the BBSRC for grant BB/P011918/1. HPA was funded through an NREL subcontract and University of Portsmouth Faculty of Science bursary. HLW thanks the US Department of Energy grant DE-SC0011297TDD. VC and ET thank the National Science Foundation grant MCB-1714556. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We thank the staff at the Diamond Light Source for their support.

Supplementary Materials and Methods

Plasmid construction

pET-21b(+) (EMD Millipore)-based plasmids for expression of the various PETase and MHETase proteins were either synthesized by Twist Bioscience or constructed using NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs) and/or the Q5® Site-Directed Mutagenesis Kit (New England Biolabs) such that each protein has a C-terminal His tag. For DNA assembly, DNA fragments were either amplified using Q5® High-Fidelity 2X Master Mix (New England Biolabs) or synthesized by Integrated DNA Technologies. Kits and master mixes were used according to the manufacturer's instructions. Plasmids were initially transformed into NEB® 5-alpha F'I^q Competent *E. coli* (New England Biolabs) and confirmed using Sanger sequencing by GENEWIZ, Inc. Specific strain construction details and sequences of synthesized DNA and PCR primers are provided in Supplementary Tables 4-1, 4-2, and 4-3, respectively.

Table 4-1: Plasmid construction.

Protein	Plasmid	Plasmid description	Construction details, reference, and other notes
PETase	pCJ135	pET-21b(+) based plasmid for expression of PETase from Ideonella sakaiensis 201-F6 (Genbank GAP38373.1), codon optimized fo rexpression in E. coli K12, with C-terminal His tag. Deposited to addgene as pET-21b(+)-Is-PETase (Plasmid 112202).	Described previously in Austin, H.P., Allen, M.D., Donohoe, B.S., Rorrer, N.A., Kearns, F.L., Silveira, R.L., Pollard, B.C., Dominick, G., Duman, R., Omari, El, K., Mykhaylyk, V., Wagner, A., Michener, W.E., Amore, A., Skaf, M.S., Crowley, M.F., Thorne, A.W., Johnson, C.W., Woodcock, H.L., McGeehan, J.E., Beckham, G.T., 2018. Characterization and engineering of a plastic-degrading aromatic polyesterase. Proc. Natl. Acad. Sci. U.S.A. 39, 201718804–8.
MHETase	pCJ136	pET-21b(+) based plasmid for expression of MHETase from Ideonella sakaiensis 201-F6 (Genbank GAP38911.1), codon optimized for expression in E. coli K12, with C-terminal His tag.	pCJ136 was constructed by assembling the DNA fragment CJ_MHETase_opt_Ec (synthesized by IDT), which ommited the stop codon to enable a C-terminal His tag, into pET-21b(+) digested with NdeI and XhoI.
01-PETlid	pCJ208	pET-21b(+) based plasmid for expression of PETase from Ideonella sakaiensis 201-F6 (Genbank GAP38373.1) incoporating the MHETase a lid, codon optimized fo rexpression in E. coli K12, with C-terminal His tag.	pCJ208 was constructed by assembling the DNA frgment CJ_MHETLid (synthesized by IDT), which ommited the stop codon to enable a C-terminal His tag, into pCJ135 digested with with NcoI and AgeI.
02-MHETnoLid	pCJ209	pET-21b(+) based plasmid for expression of MHETase from Ideonella sakaiensis 201-F6 (Genbank GAP38911.1) with the "lid" removed, codon optimized for expression in E. coli K12, with C-terminal His tag.	pCJ209 was constructed by site-directed mutagenesis of pCJ136 using NEB's Q5® Site-Directed Mutagenesis Kit according to the manufacturers intructions. pCJ136 was amplified using primer pair oCJ787/oCJ788, incorporating the lid replacement from PETase. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix.
03-MHETCC2AA	pCJ205	pET-21b(+) based plasmid for expression of MHETase from Ideonella sakaiensis 201-F6 (Genbank GAP38911.1), codon optimized for expression in E. coli K12, with C-terminal His tag, incorporating Cys224Ala and Cys529Ala mutations.	pCJ205 was constructed by site-directed mutagenesis of pCJ136 using NEB's Q5® Site-Directed Mutagenesis Kit according to the manufacturers intructions. pCJ136 was amplified using primer pair oCJ756/oCJ757 to generate a Cys224Ala mutation in the MHETase. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix. This plasmid was used as template for amplification with primer pair oCJ758/oCJ759 to generate a Cys529Ala mutation in the MHETase gene and the resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix.

04-MHETCC2WS	pCJ201	pET-21b(+) based plasmid for expression of MHETase from Ideonella sakaiensis 201-F6 (Genbank GAP38911.1), codon optimized for expression in E. coli K12, with C-terminal His tag, incorporating Cys224Trp and Cys529Ser mutations.	pCJ201 was constructed by site-directed mutagenesis of pCJ136 using NEB's Q5® Site-Directed Mutagenesis Kit according to the manufacturers intructions. pCJ136 was amplified using primer pair oCJ756/oCJ760 to generate a Cys224Trp mutation in the MHETase gene. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix. This plasmid was used as template for amplification with primer pair oCJ758/oCJ761 to generate a Cys529Ser mutation in the MHETase gene. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix.
05-MHETCC2HF	pCJ204	pET-21b(+) based plasmid for expression of MHETase from Ideonella sakaiensis 201-F6 (Genbank GAP38911.1), codon optimized for expression in E. coli K12, with C-terminal His tag, incorporating Cys224His and Cys529Phe mutations.	pCJ204 was constructed by site-directed mutagenesis of pCJ136 using NEB's Q5® Site-Directed Mutagenesis Kit according to the manufacturers intructions. pCJ136 was amplified using primer pair oCJ756/oCJ762 to generate a Cys224His mutation in the MHETase gene. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix. This plasmid was used as template for amplification with primer pair oCJ758/oCJ763 to generate a Cys529Phe mutation in the MHETase gene. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix.
06-PETCC	pCJ202	pET-21b(+) based plasmid for expression of PETase from Ideonella sakaiensis 201-F6 (Genbank GAP38373.1), codon optimized for expression in E. coli K12, with C-terminal His tag, incorporating Trp159Cys and Ser238Cys mutations.	pCJ202 was constructed by site-directed mutagenesis of pCJ135 using NEB's Q5® Site-Directed Mutagenesis Kit according to the manufacturers intructions. pCJ135 was amplified using primer pair oCJ764/oCJ765 to generate a Trp159Cys mutation in the PETase gene. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix. This plasmid was used as template for amplification with primer pair oCJ766/oCJ767 to generate a Ser238Cys mutation in the PETase gene. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix.

07-MHETCat	pCJ196	pET-21b(+) based plasmid for expression of MHETase from Ideonella sakaiensis 201-F6 (Genbank GAP38911.1), codon optimized for expression in E. coli K12, with C-terminal His tag, incorporating catalytic mutation, Ser225Ala.	pCJ196 was constructed by site-directed mutagenesis of pCJ136 using NEB's Q5® Site-Directed Mutagenesis Kit according to the manufacturers intructions. pCJ136 was amplified using primer pair oCJ768/oCJ769 to generate a Ser225Ala mutation. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix.
08-COMOSig	pCJ199	pET-21b(+) based plasmid for expression of the putative MHETase from *Comomonas thiooxydans* (Genbank WP_080747404.1) with signal peptide and C-terminal His tag, codon optimized for expression in E. coli K12.	pCJ199 was synthesized by Twist Bioscience.
09-COMOnosig	pCJ203	pET-21b(+) based plasmid for expression of the putative MHETase from Comomonas thiooxydans (Genbank WP_080747404.1) without signal peptide, with C-terminal His tag, codon optimized for expression in E. coli K12.	pCJ203 was constructed by removing the signal peptide from pCJ199 using NEB's Q5® Site-Directed Mutagenesis Kit according to the manufacturers intructions. pCJ199 was amplified with oCJ770/oCJ771 to exclude the signal peptide. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix.
10PHAGASig	pCJ207	pET-21b(+) based plasmid for expression of the putative MHETase from *Hydrogenaphaga* sp. PML113 (Genbank WP_083293388.1) with signal peptide and C-terminal His tag, codon optimized for expression in E. coli K12.	pCJ207 was synthesized by Twist Bioscience.

11-PHAGAnosig	pCJ211	pET-21b(+) based plasmid for expression of the putative MHETase from *Hydrogenaphaga* sp. PML113 (Genbank WP_083293388.1) without signal peptide, with C-terminal His tag, codon optimized for expression in E. coli K12.	pCJ211 was constructed by removing the signal peptide from pCJ207 using NEB's Q5® Site-Directed Mutagenesis Kit according to the manufacturers intructions. pCJ207 was amplided using primer pair oCJ771/oCJ772 to exclude the signal peptide. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix.
12-MHETnoLidCCwtPET	pCJ220	pET-21b(+) based plasmid for expression of MHETase from Ideonella sakaiensis 201-F6 (Genbank GAP38911.1), codon optimized for expression in E. coli K12, with C-terminal His tag, incorporating Cys224Trp and Cys529Ser mutations with lid replacement from PETase.	pCJ220 was constructed by site-directed mutagenesis of pCJ201 using NEB's Q5® Site-Directed Mutagenesis Kit according to the manufacturers intructions. pCJ201 was amplified with primer pair oCJ788/oCJ787 on 1/24/19 to generate a lid replacement from PETase. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix.
13-MHETnoLidCCdmPET	pCJ221	pET-21b(+) based plasmid for expression of MHETase from Ideonella sakaiensis 201-F6 (Genbank GAP38911.1), codon optimized for expression in E. coli K12, with C-terminal His tag, incorporating Cys224His and Cys529Phe mutations with lid replacement from PETase.	pCJ221 was constructed by site-directed mutagenesis of pCJ204 using NEB's Q5® Site-Directed Mutagenesis Kit according to the manufacturers intructions. pCJ204 was amplified with primer pair oCJ788/oCJ787 on 1/24/19 to generate a lid replacement from wtPETase. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix.
14-MHET131Gly	pCJ197	pET-21b(+) based plasmid for expression of MHETase from Ideonella sakaiensis 201-F6 (Genbank GAP38911.1), codon optimized for expression in E. coli K12, with C-terminal His tag, incorporating Ser131Gly mutation.	pCJ197 was constructed by site-directed mutagenesis of pCJ136 using NEB's Q5® Site-Directed Mutagenesis Kit according to the manufacturers intructions. pCJ136 was amplified using primer pair oCJ773/oCJ774 to generate a Ser131Gly mutation in the MHETase gene. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix.
15-MHET495Ile	pCJ198	pET-21b(+) based plasmid for expression of MHETase from Ideonella sakaiensis 201-F6 (Genbank GAP38911.1), codon optimized for expression in E. coli K12, with C-terminal His tag, incorporating Phe495Ile mutation.	pCJ198 was constructed by site-directed mutagenesis of pCJ136 using NEB's Q5® Site-Directed Mutagenesis Kit according to the manufacturers intructions. pCJ136 was amplified using primer pair oCJ775/oCJ776 to generate a Phe495Ile mutation in the MHETase gene. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix.

16-MHET6CCFae	pCJ200	pET-21b(+) based plasmid for expression of MHETase from Ideonella sakaiensis 201-F6 (Genbank GAP38911.1), codon optimized for expression in E. coli K12, with C-terminal His tag, incorporating mutations to introduce a 6th disulfide bond as in AoFaeB-2.	pCJ200 was synthesized by Twist Bioscience.
17-MHETlip	pCJ206	pET-21b(+) based plasmid for expression of MHETase from Ideonella sakaiensis 201-F6 (Genbank GAP38911.1), codon optimized for expression in E. coli K12, with C-terminal His tag, incorporating the Glu226Thr mutation to the putative lipase box.	pCJ206 was constructed by site-directed mutagenesis of pCJ136 using NEB's Q5® Site-Directed Mutagenesis Kit according to the manufacturers intructions. pCJ136 was amplified using primer pair oCJ777/oCJ778 to generate a Glu226Thr mutation in the MHETase gene. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix.
18-MHET6CCPET	pCJ217	pET-21b(+) based plasmid for expression of MHETase from Ideonella sakaiensis 201-F6 (Genbank GAP38911.1), codon optimized for expression in E. coli K12, with C-terminal His tag, incorporating two point mutations, Gly489Cys and Ser530Cys, to introduce a 6th disulfide bond (from PETase).	pCJ217 was constructed by site-directed mutagenesis of pCJ136 using NEB's Q5® Site-Directed Mutagenesis Kit according to the manufacturers intructions. pCJ136 was amplified using primer pair oCJ779/oCJ780 to generate Gly489Cys mutation in the MHETase gene. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix. This plasmid was used as template for amplification with primer pair oCJ781/oCJ782 to generate a Ser530Cys mutation in the MHETase gene. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix.
19- MHET7CC	pCJ210	pET-21b(+) based plasmid for expression of MHETase from Ideonella sakaiensis 201-F6 (Genbank GAP38911.1), codon optimized for expression in E. coli K12, with C-terminal His tag, incorporating mutations to introduce a 6th and 7th disulfide bond as in AoFaeB-2.	pCJ210 was constructed by site-directed mutagenesis of pCJ200 using NEB's Q5® Site-Directed Mutagenesis Kit according to the manufacturers intructions. pCJ200 was amplified using primer pair oCJ779/oCJ780 to generate Gly489Cys mutation in the MHETase gene. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix. This plasmid was used as template for amplification with primer pair oCJ781/oCJ782 to generate a Ser530Cys mutation in the MHETase gene. The resulting PCR product was treated with NEB's Kinase, Ligase, and DpnI (KLD) enzyme mix.

Table 4-2: Synthesized DNA fragments.

Fragment	Sequence (5'-3')	Description	Length (bp)
CJ_MHETase_opt_Ec	ctttaagaaggagatataCAT**ATG**cagaccaccgtgac caccatgctcctcgcgtccgtagcattagcggcttgcgccg gaggaggttccactcctctgcctctaccgcagcagcagccg cctcagcaggaaccgccacctcctcctgttccgctagccagt cgcgccgcgtgtgaggcgctcaaagatggtaatggcgaca tggtttggccgaatgccgccacggttgtagaggttgcagcct ggcgtgatgcagcaccggccacggcatcagccgcagccc tgccggagcattgcgaagtatcaggcgcgattgccaagcgt actgggattgatgggtacccgtatgaaattaagtttcgcctgc gcatgcccgctgagtggaacggccgttttttcatggagggtg gcagtggtacgaacggctctctctcagcggcgaccggaag tatcggcggcggtcagatcgcctcagcgctgagtcgtaactt tgcaacaattgctaccgacggaggacatgacaatgcggtga atgataatccggatgcgctcggtaccgtcgcatttggtctcga tccccaggcacgcttagacatgggctacaactcctatgatca ggtgactcaggccggcaaagccgccgttgcacgcttttatg gtcgcgcagccgacaagagctacttcatcggctgttcggag ggcggccgcgagggcatgatgctgtcccagcgctttccatc acattacgatggcattgtggcgggcgcaccgggatatcagtt gccgaaggccggaattagtggcgcgtggaccacccagag cttagcgcccgccgccgttggcctggatgcccagggagtg ccgctgattaataagagcttttctgacgcagacctccatttact gtcgcaggcgattctcggaacatgcgacgccttggatggcc tggccgacggcatcgttgacaactaccgagcgtgccaagc ggcttttgatccggcgactgcagccaacccagcgaatggcc aagccctgcagtgcgtgggcgcaaagacagccgattgctta tcgcccgtccaagttacggcgattaaacgagcgatggccgg tccggtaaatagcgcgggtacgccgttatataatagatgggc ctgggacgcaggtatgagcggtcttagtggtaccacttacaa tcagggttggcgcagctggtggctgggatcgtttaacagctc ggcgaataacgcacaacgtgtatctggtttctcagcgcgga gctggctggtggactttgctaccccgccggagccgatgccc atgacccaagtcgccgcccgtatgatgaaatttgatttcgata tcgatcctctgaaaatatgggctacttcgggccaatttaccca gagtagtatggactggcacggtgccactagcaccgaccttg ctgcctttcgggaccgcggcggtaaaatgattctgtatcacg gaatgagcgatgccgcattctctgcactagatacagcagatt attatgaacgcctgggtgccgcaatgccgggcgccgcgggg ctttgctcgtctgttcttggttccgggaatgaaccattgctccg ggggtccaggtaccgaccgctttgatatgctaacaccgttag ttgcatgggttgaacgtggggaagcccctgaccaaattagc gcctggagcggcacccccggctactttggtgtggccgccc gcactcgaccgttatgtccctatccgcagattgcgcgctataa gggatcaggcgatatcaataccgaagcaaattttgcgtgtgc cgctccaccg<u>CTCGAGcaccaccatcaccaccac**tga**</u> <u>gatccggct</u>	The MHETase from Ideonella sakaiensis strain 201-F6 (Genbank GAP38911.1) was codon optimized for expression in E. coli K12 MG1655 (Highly Expressed Genes) using the codon optimizer at http://genomes.urv.es/OPTIMIZER/ (guided random method) (CAI: 0.658, ENc: 52, %GC: 58.5). The stop codon was ommited and overlaps were added for assembly into the expression vector pET-21b(+) digested with NdeI and XhoI such that the assemply would results in the addition of a C-terminal 6 His tag. The 6 His sequecnce in pET-21b(+) is the same codon (CAC) repeated 6 times and so one of the His codons in the synthesized DNA fragment was changed to CAT to enable assembly.	1866

CJ_MHETase_Lid	tccgcgcttagctagccatggctttgtggttattaccatcgata cgaacagcactctagaccagcccagcagccgtagctcgca acagatggccgcgcttcgtcaagttgcgagcttgaacggga ccagcagtagcccgatttacggaaaggtcgatactgcccgc atgggtgtgatgggctggtcaatgggggcggcggttcact tattagcgccgcgaacaacccgagtttaaaagcagcggcac cgcaggcgccaGGATATCAGTTGCCGAAG GCCGGAATTAGTGGCGCGTGGACCA CCCAGAGCTTAGCGCCCGCCGCCGTT GGCCTGGATGCCCAGGGAGTGCCGC TGATTAATAAGAGCTTTTCTGACGCA GACCTCCATTTACTGTCGCAGGCGAT TCTCGGAACATGCGACGCCTTGGATG GCCTGGCCGACGGCATCGTTGACAA CTACCGAGCGTGCCAAGCGGCTTTTG ATCCGGCGACTGCAGCCAACCCAGC GAATGGCCAAGCCCTGCAGTGCGTG GGCGCAAAGACAGCCGATTGCTTAT CGCCCGTCCAAGTTACGGCGATTAAA CGAGCGATGGCCGGTCCGGTAAATA GCGCGGGTACGCCGTTATATAATAG ATGGGCCTGGGACGCAGGTATGAGC GGTCTTAGTGGTACCACTTACAATCA GGGTTGGCGCAGCTGGTGGCTGGGA TCGTTTAACAGCTCGGCGAATAACGC ACAACGTGTATCTGGTTTCTCAGCGC GGAGCTGGCTGGTGGACTTTGCTACC CCGCCGGAGCCGATGCCCATGACCC AAGTCGCCGCCCGTATGATGAAATTT GATTTCGATATCGATCCTCTGAAAAT ATGGGCTACTTCGGGCCAATTTACCC AGAGTAGTATGGACTGGCACGGTGC CACTAGCACCagcagtgttaccgtgccgacgctga ttttcgcgtgcgagaatgatagcattgcaccggtgaacagca gcgcgct	"Lid" region from the Ideonella sakaiensis strain 201-F6 (Genbank GAP38911.1) MHETase, with overlaps for assembly into pCJ135. Assembly will generate PETase with the MHETase lid.	999

pCJ199	CAT**ATG**ttcgtacgcaacgccgaccgtgccaagaatt gtatgcgcgcacctttacgcgtattcccactcaaggatacttttt agcgcccagtgtgcgaatgtttcggtctggattaccagcagc gtaccaccgctccgcgagcgtcacatggatcgccgcgtga cgcgccgcgatttaatgcaaactcgcatcttattaatgctcatt gcagccactggcgtggcagcgtgtggcggagacggtggtt ccacacctgccgcgcaaaatccccctttgcccctggccagt cgtgcggcttgcgaagcttttcaaggcaatagcaatagtatc gcgtggccccatcgcgcaaccgttgtggaagtggccacttg gcacgaagcagagcctgcgaatgccacagcagcggcgac gcccgagcactgtgagatttccggcgccattgctcgccgca ccggaattgatggatatccttacgagattaagtttcgcttacgt atgccctcagaatggaatggtcgcttctttatggaaggggggg ggtggaaccaatgggtcattgagtgccgctacagggtccctt ggcggtggacaaactgcgtcggccttgagtcgtaattttgca actattgcaaccgatggtggtcatgataatgctgtcaacaata atcctgatgcgctcggcactgtcgcttttggcatggaccctca agcgcgcattgatatgggatataattcctacgaccaggtgac ccaagcgggaaaggcggccgtagcgcaattttatggccgt gccccggataaaagctattttattggctgtagcgaaggtggg cgcgaggggatgatgttgtcccaacgctttccgagtcattat gacggaattgtggcgggagcaccgggctaccaactcccaa aggcgggcatttctggcgcatggacgacgcaaagtctggc accagcagcggtgggcgtggatgctcaaaatgtacctttgat caataaggcgttttcggatgtcgatttacatcttctttcacgcg gcattcttggtacttgcgatgccttggatggactcagcgatgg aattgtgaacgacttccgtgcctgtcaagccgcctttgaccct gccactgcgttgaatcccgacaccagtcaaccccttacaatgc actggtgctaagacgcctgattgcttaagtgccgcccaagtc actggcattaaacgcgccatgggtgggcctgtggacagcg ccggtgcggcattatataatcgctgggcatgggaccctggc atgtcggggctcaatggcacctcttacaatcagggatggcg ctcttggtggttagggagctacgcatccagcactaataatgc ccaacgcgtcaatggcttttccgcccgctcttggttggttgatt ttgcgacgccgcctgaaccaatgccagtcacacaggttgct gctcgcatgatgaacttcaactttgacaccgacccgcctaag attcgcgcgactagtggccccttactccatcgtctatggagt ggcatggtgcaacgagcactaatcttgcggccttccgcgac cgcggtgggaagctgatgctctatcatggcatgtcagatgcc gcgttctccgcattagacaccgcggattactacgagcgttta ggtgccgcgatgccgggcgccgccggcttcgcacgtctttt cttagttccaggtatgaatcattgtagtggtggacccggcact gatcgttttgatatgcttactccccttgtggcctgggtggaacg tgataaggcgccagatcaagttagcgcttgggcaggcacac cgggctatttcggcgcaaccgcccgtacacgccccctttgtc catacccccaaatcgcacgttataagggctctggtgatatca atgccgaggcatcgtttgtgtgtggccccaCTCGAG	pCJ199 was synthesized by Twist Bioscience. Only the sequence integrated between the NdeI and XhoI sites in pET-21b(+) is shown.	

pCJ200	CAT**ATG**cagaccaccgtgaccaccatgctcctcgcgt ccgtagcattagcggcttgcgccggaggaggttccactcct ctgcctctaccgcagcagcagccgcctcagcaggaaccgc cacctcctcctgttccgctagccagtcgcgccgcgtgtgag gcgctcaaagatggtaatggcgacatggtttggccgaatgc cgccacggttgtagaggttgcagcctacgtgccggcaggc gttaacatcagcatggcggataacccgagcatctgtggtgg cgacgaggacccgattacttccaccttcgcgttctgcgaagt atcaggcgcgattgccaagcgtactgggattgatgggtacc cgtatgaaattaagtttcgcctgcgcatgcccgctgagtgga acggccgttttttcatggagggtggcagtggtacgaacggct gcctctcagcggcgaccggaagtatcggcggcggtcagat cgcctcagcgctgagtcgtaactttgcaacaattgctaccga cggaggacatgacaatgcggtgaatgataatccggatgcg ctcggtaccgtcgcatttggtctcgatccccaggcacgctta gacatgggctacaactcctatgatcaggtgactcaggccgg caaagccgccgttgcacgcttttatggtcgcgcagccgaca agagctacttcatcggctgttcggagggcggccgcgaggg catgatgctgtcccagcgctttccatcacattacgatggcatt gtggcgggcgcaccgggatatcagttgccgaaggccgga attagtggcgcgtggaccacccagagcttagcgcccgccg ccgttggcctggatgcccagggagtgccgctgattaataag agcttttctgacgcagacctccatttactgtcgcaggcgattct cggaacatgcgacgccttggatggcctggccgacggcatc gttgacaactaccgagcgtgccaagcggcttttgatccggc gactgcagccaacccagcgaatggccaagccctgcagtgc gtgggcgcaaagacagccgattgcttatcgcccgtccaagt tacggcgattaaacgagcgatggccggtccggtaaatagc gcgggtacgccgttatataatagatgggcctgggacgcagg tatgagcggtcttagtggtaccacttacaatcagggttggcg cagctggtggctgggatcgtttaacagctcggcgaataacg cacaacgtgtatctggtttctcagcgcggagctggctggtgg actttgctaccccgccggagccgatgcccatgacccaagtc gccgcccgtatgatgaaatttgatttcgatatcgatcctctgaa aatatgggctacttcgggccaatttacccagagtagtatgga ctggcacggtgccactagcaccgaccttgctgcctttcggg accgcggcggtaaaatgattctgtatcacggaatgagcgat gccgcattctctgcactagatacagcagattattatgaacgcc tgggtgccgcaatgccgggcgccgcgggctttgctcgtctg ttcttggttccgggaatgaaccattgctccgggggtccaggt accgaccgctttgatatgctaacaccgttagttgcatgggttg aacgtggggaagcccctgaccaaattagcgcctggagcgg caccccccggctactttggtgtggccgcccgcactcgaccgtt atgtccctatccgcagattgcgcgctataagggatcaggcg atatcaataccgaagcaaattttgcgtgtgccgctccaccgC TCGAG	pCJ200 was synthesized by Twist Bioscience. Only the sequence integrated between the NdeI and XhoI sites in pET-21b(+) is shown.	

pCJ207	CAT**ATG**aagtctagcattccgattagcgtaggtatgctg gctaccgctctgatttccggttgcggtagcgttccggataaca ccagcaatgaaccgacggttccgctggcatccaaagaattg tgcgagggcatcgcgtctggtgcgaccaaagtaaactggc cgaaccagaacaccgtcgtaaaagcttcagtttggcacgct gttaccccggcaaccgccaacgccccggaactgccggaa cattgcgaggtcactggctctatcaaccagcgtactggcgtg gacggctatccgtatgaaatcaaaatgcgtctgcgcatgccg gcagattggaacggccgtttcttcatggaaggcggtggagg tactaacgggagcctgtctgccgctctgggttcgctgggcg gtggtcagaccagcaatgctctgagccgtcgtttcgctaccg tttctaccgatggtggtcatgataacgcagtgaacaacaatcc ggcggcgctgggttcggtcgctttcggcatggacccgcag gctcgcctggatcatggttacaactcatacgatcaggttaccc tggcgggcaagtcagcagtaagcacttttttacgggcgcggc ccggacaaatcatacttcatcggctgttccgaaggcggtcgt gagggcatgatgttcagccagcgtttcccggcgcactatga cggcatcgtcgccggtgcaccgggctaccagctgccgaaa gcaggcatcagcggtgcatggacaacgcaatctctggcac cagcggccgttggtgttgacccggacggtgcaccgctggt gaacaaatccttcagcgatccggacctgtatctgctgactca ggcaatcctgggcagctgcgacgctctggacggtctggctg acggtatcgtcggcaattattccgcgtgtcagtcgctgtttga cccgtctaccgccttgaaccctgcgacgggacaaccactgc gttgcacgggcgctaagaccgacgactgcctaagcccggtt caggtggatgcgatcaaacgtgctatgtccggtccagttgat actgccggaaccgccctgtataacaaatggccgtgggatac cggtatgtcgggcctgaacggcaccacttatttccagggctg gcgcagctggtggctgggctcctacgacagctctactaaca acgcgcagcgtgttaacggcagcagcgcacgctcttggct ggtagatttcgctactccgccggaacctgtaccgctgaacca ggtggccactcgtatgatgaactttgattttgatgttgacccgc cgaaaatctttgctacctctggtctctttacccagccgtccatg caatggcacggtgccacctcaaccgatctgaacgcttttcgc tctcgcggtggcaagctgatgctgtaccacggcatggctga cgcggcattcagcgcactggataccattgcttattatgagcg cctgagcaccgcaatgccttccgtgtccgacttttctcgcctg tttctggtgcctggtatggggcactgttccggcggtccgggc accgatcgctttgatatgctgactccgctggtggcgtgggttg agaacggtactgcaccggctcgcgtcgaagcgtcgtcctcc actccgggttacttcggtgtttcggcccgcagccgccccctg tgcccgcatccgcagattgcacgttataccgggtccggcga cattaacgaagccaccaactttgtatgcggtaacccgCTC GAG	pCJ207 was synthesized by Twist Bioscience. Only the sequence integrated between the NdeI and XhoI sites in pET-21b(+) is shown.	

Table 4-3: PCR primers.

Oligo	Sequence (5' -> 3')	Description
oCJ756	**tcggagggcggccg**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Ser225 F
oCJ757	GGC**gccgatgaagtagctcttgtcggc**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Ser225 R with Cys224Ala mutation
oCJ758	**tccgggggtccaggtacc**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Ser530 F
oCJ759	GGC**atggttcattcccggaaccaagaacag**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Ser530 R with Cys529Ala mutation
oCJ760	CCA**gccgatgaagtagctcttgtcggc**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Ser225 R with Cys224Trp mutation
oCJ761	GCT**atggttcattcccggaaccaagaacag**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Ser530 R with Cys529Ser mutation
oCJ762	GTG**gccgatgaagtagctcttgtcggc**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Ser225 R with Cys224His mutation
oCJ763	GAA**atggttcattcccggaaccaagaacag**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Ser530 R with Cys529Phe mutation
oCJ764	**tcaatgggggggcggcg**	For linear amplification of Ideonella sakaiensis PETase expression plasmid, pCJ135, at Ser160 F
oCJ765	GCA**gcccatcacacccatgcgg**	For linear amplification of Ideonella sakaiensis PETase expression plasmid, pCJ135, at Ser160 R with Tryp159Cys mutation
oCJ766	**tgtgccaactctgggaacagc**	For linear amplification of Ideonella sakaiensis PETase expression plasmid, pCJ135, at Cys239 F
oCJ767	GCA**gtggctaccgccgttaatttccag**	For linear amplification of Ideonella sakaiensis PETase expression plasmid, pCJ135, at Cys239 R with Ser238Cys mutation
oCJ768	**gagggcggccgcga**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Glu226 F
oCJ769	GGC**acagccgatgaagtagctcttgtcg**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Glu226 R with Ser225Ala mutation
oCJ770	**TGTGGCGGAGACGGTG**	For linear amplification of Comomonas thiooxydans putative MHETase plasmid, pCJ199, at Cys76 F
oCJ771	**CATATGtatatctccttcttaaagttaaacaaaatta tttcta**	For linear amplification of putative Comomonas thiooxydans MHETase expression plasmid, pCJ199, and putative MHETase from Hydrogenaphaga sp. expression plasmid, pCJ211, at Met1 R
oCJ772	**TGCGGTAGCGTTCCGG**	For linear amplification of putative MHETase from Hydrogenaphaga sp. expression plasmid, pCJ211, at Cys20 F
oCJ773	GGC**ggtacgaacggctctctctcag**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Ser131 F with Ser131Gly mutation
oCJ774	**gccaccctccatgaaaaaacgg**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Ser131 R
oCJ775	ATC**tctgcactagatacagcagattattatgaac**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Phe495 F with Phe495Ile mutation
oCJ776	**tgcggcatcgctcattcc**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Phe495 R

oCJ777	**ggcggccgcgagg**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Gly227 F
oCJ778	GGT**cgaacagccgatgaagtagctcttgtc**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Gly227 R with Glu226Thr mutation
oCJ779	TGC**atgagcgatgccgcattctctg**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Gly489 F with Gly489Cys mutation
oCJ780	**gtgatacagaatcattttaccgccgcg**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Gly489 R
oCJ781	**gggggtccaggtaccgac**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Ser530 F
oCJ782	GCA**gcaatggttcattcccggaacc**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, at Ser530 R with Ser530Cys mutation
oCJ783	gtttaaaagcagcggcaccgcaggcgcca**GGATATCAGTTGCCGAAGGCCG**	For amplification of Ideonella sakaiensis MHETase lid domain F with oCJ784 overlap
oCJ784	**ccgctgcttttaaactcgggttgttc**	For linear amplification of Ideonella sakaiensis PETase expression plasmid, pCJ135, R with oCJ783 overlap
oCJ785	acactgct**GGTGCTAGTGGCACCGTGC**	For amplification of Ideonella sakaiensis MHETase lid domain R with Ideonella sakaiensis PETase expression plasmid, pCJ135, overlap
oCJ786	TAGCACC**agcagtgttaccgtgccgac**	For linear amplification of Ideonella sakaiensis PETase expression plasmid, pCJ135, F with oCJ785 overlap
oCJ787	CAACCAACTTC**gaccttgctgcctttcgggac**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, F
oCJ788	AAGAGTCCCA**cggtgcgcccgcc**	For linear amplification of Ideonella sakaiensis MHETase expression plasmid, pCJ136, R

Ligand synthesis

MHET and iMHET synthesis. Mono (hydroxy 2-ethyl) terephthalate (MHET) and mono(hydroxy 2-ethyl) isophthalate (iMHET, also referred to as MHEI) were synthesized via the condensation of either terephthloyl chloride or isophthloyl chloride with monoprotected ethylene glycol (tBOC-EG) which was subsequently deprotected to yield the final product.

Initially, tBOC-EG was prepared by stirring one (1) molar equivalent (a large excess) of ethylene glycol (EG) with 0.0575 equivalents of di-tert-butyl decarbonate with 0.0051 equivalents of 4-dimethylaminopyridine (DMAP) as a catalyst in dichloromethane (DCM). The reaction mixture was allowed to stir for 24 hours and was subsequently washed with DI water, 1M HCl, and brine follow by drying with sodium sulfate. The yield of this reaction was 20% at a final purity of 99+% (via NMR and HPLC). The solvent was then removed, and the product was used "as-is" in subsequent reaction steps.

To form MHET or iMHET either one (1) molar equivalent of terephthloyl chloride or isophthloyl chloride respectively was dissolved in DCM with one (1) molar equivalent of tBOC-EG. One molar equivalent of triethylamine (TEA) was then added in dropwise over a period of 30 minutes. The reaction solution was subsequently washed with DI water and brine and then dried over sodium sulfate prior to removing the DCM. The crude product was subsequently taken up in a mixture of 10% acetone in DCM and purified via silica gel chromatography (Figure 4-5).

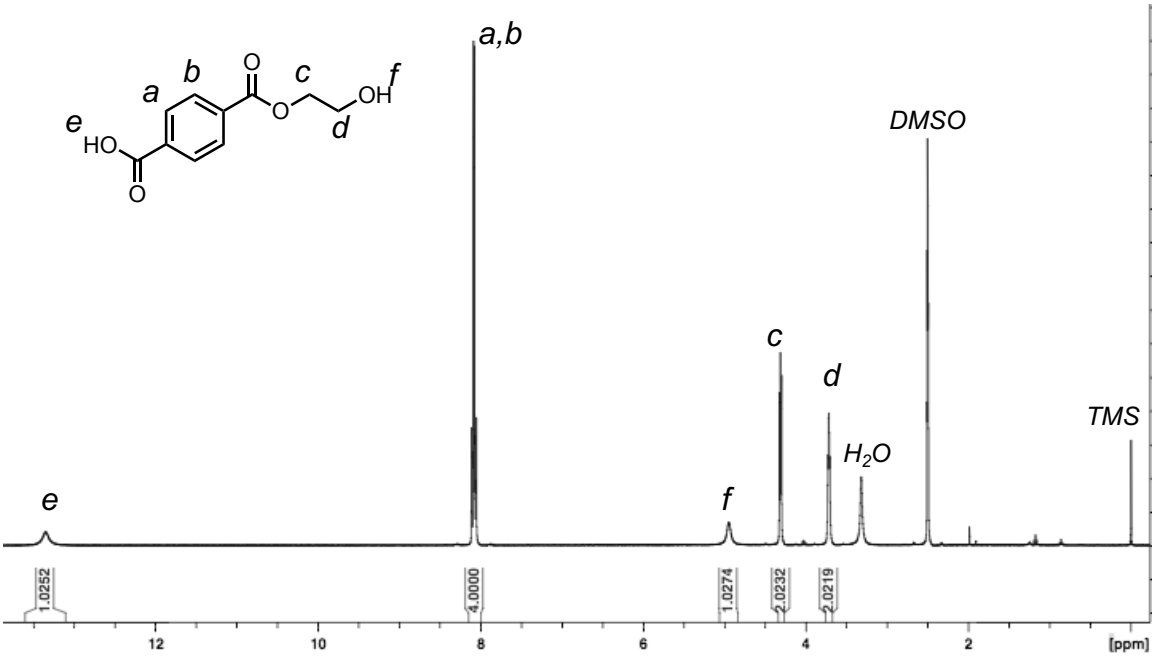


Figure 4-5: ^{1}H NMR of MHET with peak assignments. Integration and peak splitting confirm MHET was formed and that the product is not BHET.

Crystallography

Native MHETase was expressed in *E. coli* C41(DE3) in 2xTY media and protein production induced by the addition of IPTG (1 mM final concentration). The cells were harvested by centrifugation, disrupted by sonication and the cell debris removed by centrifugation at 20,000 rpm for 30 minutes. The protein purified by Ni-affinity chromatography followed by gel filtration on a Superdex 75 HR column. The protein was concentrated to a range of concentrations (9-14 mg/mL) and dialyzed into 10 mM Tris, pH 7.0, 100 mM NaCl for crystallography. K-MOPS minimal media [137] was used for seleno-methionine labeling of MHETase. Cells were grown to an OD600 of 0.5 after which 100 mg/L of DL-seleno-methionine (Sigma), 100 mg/L lysine, threonine and phenylalanine, leucine, isoleucine and valine were added as solids. IPTG (1 mM final

concentration) was then added after a further 20 min and cells were grown for a further 16 h at 20°C. Seleno-methionine labeled protein was purified as described above.

MHETase was crystallized at a range of concentration from 9-14 mg/mL by sitting-drop vapor diffusion. Several conditions yielded crystal, four of which were used to obtain datasets, one of which contained seleno-methionine labeled protein.
The crystals were flash-frozen in liquid nitrogen after the addition of glycerol to 20% while leaving the other components of the mother liquor at the same concentration. Seleno-methionine MHETase crystals belonged to space group $P2_12_12$ were used to obtain phase information using the I03 beamline at Diamond. Data were obtained from 3600 images collected at 0.9795 Å with 0.1° increments. All images were integrated using XDS [138]and scaled using SCALA [139]. Phases were obtained using PHASERSAD in the ccp4i software in combination with PARROT and SHELXD [140,141].The initial output was subsequently built using BUCCANEER and further refined using iterative rounds of COOT and PHENIX[142–144]. One molecule of MHETase was observed in the asymmetric unit of the $P2_12_12_1$ SeMet SAD dataset. Three additional native datasets, each containing 1800 images collected at 0.1° increments, were collected at beamline I03 of the Diamond Light Source. The structure of native MHETase were obtained using molecular replacement from a refined molecule of MHETase obtained initially from the SeMet-SAD data. All structures were refined using iterative rounds of COOT and PHENIX. Cell constants and crystallographic data are summarized in Table 4-1. Details of the refined models are shown in Table 4-1. Figures were generated with PYMOL.

Table 4-4: Data collection and structure refinement of MHETase.

Data set	SeMet	Native 1	Native 2	Native 3
Symmetry	$P2_12_12$	$P2_12_12$	$P2_12_12_1$	P1
Wavelength	0.9795	0.9795	0.9795	0.9795
Resolution Range (Å)	∑	46.00 – 1.70	46.24 – 1.80	95.29 - 1.90
Unique reflections	84000	70855	122377	506864
Completeness (%)[a]	99.8 (98.9)	99.6 (98.0)	98.2 (96.9)	93.7 (92.4)
Anomalous Completeness (%)[a]	99.8 (98.5)			
*R*merge[b]	0.063 (0.511)	0.056 (0.584)	0.067 (0.550)	0.057 (0.254)
CC(1/2)[c]	0.999 (0.902)	0.999 (0.900)	0.999 (0.914)	
Multiplicity[d]	12.6 (11.0)	6.4 (6.5)	6.7 (6.6)	1.7 (1.7)
Anomalous Multiplicity[d]	6.6 (5.6)			
I/σI[a]	20.2 (4.6)	15.4 (3.1)	14.7 (2.9)	7.1 (2.6)
	a = 77.37 Å, b = 89.02 Å, c = 91.64 Å	a = 77.20 Å, b = 89.88 Å, c = 92.00 Å	a = 90.21 Å, b = 92.80 Å, c = 159.99 Å	a = 110.49 Å, b = 135.63 Å, c = 138.15 Å, α = 83.09°, β = 67.91°, γ = 67.57°
Model Refinement				
Resolution Range (Å)	45.82 - 1.60	46.00 – 1.60	46.24 - 1.80	46.46 – 1.90
No. of residues:	A: 40-55, 62-603	A: 40-55, 61-603	A: 36-603, B: 36-603	A: 42-603, B:43-603, C: 43-603, D: 43-603, E: 43-603, F: 43-603, G: 41-603, H: 42-603, I: 43-603, J: 43-603
No. of water, ligands	748, 1 Ca, 1 benzoic acid	536, 1 Ca, 1 benzoic acid	1407, 2 Ca	6125, 10 Ca
Rwork/Rfree (%)[e]	16.20 (17.80)	16.55 (19.21)	18.24 (20.51)	18.54 (20.54)
B average[f]	25.8	32.8	30.9	28.8
Geometry bond, angles[g]	0.003, 0.613	0.009, 0.975	0.005, 0.715	0.003, 0.576
Ramachandran[h]	97.47, 0.0	97.48, 0.2	97.0, 0.0	97.07, 0.02
Molprobity Clash Score	1.61	1.48	3.70	4.15
Beamline	I03	I03	I03	I03
PDB ID[i]	6QZ3	6QZ1	6QZ4	6QZ2

Table 4-4 (Continued): [a] Signal to noise ratio of intensities, highest resolution bin in brackets. [b] R_m : $\sum h \sum i |I/(h,i) - I(h)| / \sum h \sum i / I(h,i)$ where $I(h,i)$ are symmetry-related intensities and $I(h)$ is the mean intensity of the reflection with unique index h . [c] $CC_{1/2}$ is the correlation coefficient of the mean intensities between two random half-datasets. [d] Multiplicity for unique reflections. [e] 5% of reflections were randomly selected for determination of the free R factor, prior to any refinement. [f] Temperature factors averaged for all atoms. [g] RMS deviations from ideal geometry for bond lengths and restraint angles [145h]. Percentage of residues in the 'most favored region' of the Ramachandran plot and percentage of outliers (MOLPROBITY). [i] Protein Data Bank identifiers for coordinates.
Crystallography conditions: SeMet; 0.1 M sodium cacodylate (pH 6.5), 9% PEG 8000. Native 1; 0.1 M sodium acetate (pH 5.5), 24% PEG 5000 MME. Native 2; ammonium acetate (pH 4.5), 22.5% PEG 10000. Native 3; 0.1 M sodium citrate (pH 5.5), 1.0 M ammonium phosphate monobasic. All conditions used 20% glycerol as a cryoprotectant.

Expression and purification of
PET mutant, MHETase mutants and putative MHETases

After initial screening for soluble protein expression for the proteins of interest, various cell lines and induction methods[88] were used to purify protein for kinetic assays. Types of cell lines used for each protein is described in Table 4-6. Generally, cells were transformed with a specific plasmid and single colonies from transformants were inoculated in a starter culture containing either 100 μg/mL ampicillin or 100 μg/mL ampicillin and 34 μg/mL chloramphenicol. After the protein induction, cells were harvested by centrifugation, and stored at -80 until further use. Harvested cells were resuspended in a lysis buffer (400 mM NaCl, 20 mM Tris-Cl pH 8.0) and lysed using a microfluidizer. Lysate was clarified via centrifugation at 20,000 x g for 30-45 minutes. Supernatant was collected and then applied to a Ni-NTA gravity flow column. The nickel affinity column was washed with 3 times with 2 column volumes of wash buffer (400

mM NaCl, 20 mM Tris-Cl pH 8.0, 10 mM imidazole), and the protein fraction was eluted using elution buffer (100 mM NaCl, 20 mM Tris-Cl pH 8.0, 300 mM imidazole), and repeating this procedure five times. Cell pellets, wash fractions and column eluents were analyzed for protein content using 1D SDS-PAGE. Protein was then dialyzed twice against 500 ml 100 mM NaCl, 20 mM Tris-Cl pH 7.0 for two hours and overnight, respectively. Afterwards, protein was concentrated using a 10,000 Daltons molecular weight cut-off (MWCO) Amicon Ultra centrifugal filter and resulting filtered solution applied to a Superdex-75 size exclusion chromatographic column equilibrated with 100 mM NaCl and 20 mM Tris-Cl pH 7.0. Protein concentrations were assessed using Bradford assays [89].

Analysis of Michaelis Menten kinetics of MHETase variants

2 nM MHETase was incubated with varying MHET concentrations ranging from 2 μM to 25 μM in 40mM phosphate buffer with 10% (v/v) DMSO at pH 7.2 at 30°C. The reaction was terminated using an equal volume of 160 mM phosphate buffer (pH 2.5). Initial rates were plotted as a function of produced TPA and kinetic parameters were determined by using double reciprocal plots.

Enzyme incubation with substrate

lcPET (3-4%, Goodfellow – USA) was incubated with enzyme of interest (PETase or MHETase) in polypropylene tubes containing 50 mM pH 7.2 phosphate buffer at 30°C for 96 hours. The reaction was terminated by heat shock at 85°C for 10 minutes. PET coupon removed from reaction was washed with 1% SDS, ultrapure water,

and 95% ethanol. Coupons were then vacuum dried overnight at 60°C for scanning electron microscopy.

Scanning Electron Microscopy

Dried PET coupons sized 2.5 cm x 0.5 cm were placed on aluminum stubs using carbon tape, and were sputter coated with 9 nm of iridium. SEM imaging was performed using an. FEI quanta 400 FEG instrument under low vacuum (0.45 torr), beam-accelerating voltage of 15 keV.

PETase and MHETase synergy

For PETase and MHETase synergy, effects of two independent factors, PETase load (mg/g PET) and MHETase load (mg/g PET) were explored as a function of mass loss of lcPET. For the initial explorative round, we have used a full factorial analysis. Each factor had enzyme load levels ranging from 0 mg/g PET to 3 mg/g PET with 0.25 mg/g PET increments in triplicates. Enzyme activity response was measured as lcPET mass loss following incubation at 30°C for 96 hours in 50 mM sodium phosphate buffer, pH 7. After the initial screening, optimal enzyme loading range of PETase and MHETase for improved mass loss was determined, a second round of optimization was conducted using surface response methodology. For the second round of optimization, enzyme loadings ranging from 0 to 0.375 mg/g PET for PETase and 0 to 0.75 mg/g PET for MHETase (in triplicates) were used. Enzyme activity response was measured as lcPET mass loss following incubation at 30°C for 96 hours in 50 mM sodium phosphate buffer, pH 7. For model generation, lcPET mass loss was used as a function of enzyme loading.

Design Expert (Version 11.0, Stat-Ease Inc., USA) was used for response surface generation, evaluation of the model and the fit. The quadratic equation used to fit the model is as follows (Table 2-5 and Table 2-6): ln(mass loss) = -3.94 + 15.54352*A + 1.05674*B + 4.81714*AB – 29.66140*A2 – 2.53290*B2. The model was statistically analyzed to evaluate the analysis of variance (ANOVA) and fit (Table 2-7). The modeled surface was used to determine the optimum PETase and MHETase load for maximum response in terms of lcPET mass loss.

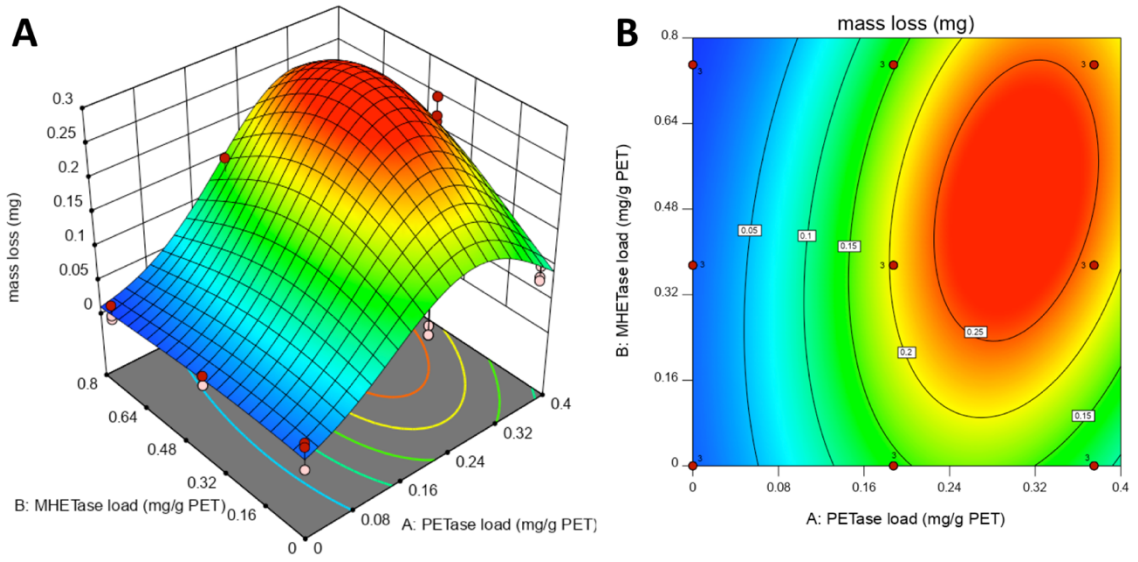


Figure 4-6: A) Modeled 3D response surface of PETase and MHETase synergy. PETase loading (mg/g PET, axis A) and MHETase loading (mg/g PET, axis B) are modeled against mass loss of lcPET after 96 hours. Data points above and below the predicted values are colored red and pink, respectively. Blue surface coloration indicates low mass loss, whereas red surface coloration indicates higher mass loss. B) The contour plot (level curves) of the response surface of mass loss.

MHETase activity on MHEI

Activity assay of MHETase with MHET, MHEI. Assay of MHETase activity was performed using 5nM enzyme concentration and 25uM, 50uM, and 250uM substrate

concentration at 30˚C for 24h in a 1mL reaction volume. The enzyme activity assays were carried out in 50mM Na-Phosphate, pH7.5, 100mM NaCl reaction buffer with three concentrations of substrate (either MHET or MHEI) in triplicate. Reactions were started by addition of enzyme or an equal volume of reaction buffer for the no enzyme controls. The no enzyme control reactions at each substrate concentration were performed in duplicate. At the end of 24h the reactions were treated with 500uL DMSO, filtered through 0.2 μm nylon membranes, and analyzed by HPLC.

Table 4-5: Fit summaries of equations of various orders tested for response surface modeling.

Source	Sequential p-value	Lack of Fit p-value	Adjusted R^2	Predicted R^2
Linear	< 0.0001	0.0013	0.6128	0.5401
2FI	0.1219	0.0017	0.6368	0.5330
Quadratic	< 0.0001	0.6057	0.8383	0.7627
Cubic	0.3880	0.9862	0.8382	0.7433

HPLC method Standards of terephthalate and isophthalate were obtained from Sigma Aldrich and the MHEI and MHET substrates were synthesized in-house at NREL . Analyte analyses were performed on an Agilent 1200 LC system (Agilent Technologies, Santa Clara, CA) equipped with a G1315A diode array detector (DAD). Each sample and standard were injected using a volume of 10 μL onto a Phenomenex Luna C18(2) column, 5 μm, 4.6 x 150 mm column (Phenomenex, Torrance, CA). The column temperature was maintained at 40°C and the buffers used to separate the analytes of

interest consisted of 20 mM Phosphate buffer in water(A) and methanol(B). The separation was carried out using a constant flow rate of 0.6mL/min and a gradient program of: at t = 0min (A) = 80% and (B) = 20%; at t = 15min (A) = 35% and (B) = 65%; at t = 15.01min through 20min (A) = 80% and (B) = 20% for a total run time of 20min. The calibration curve for each analyte was evaluated between concentrations of 1 – 200mg/L. DAD detection at a wavelength of 240nm was performed for each analyte. A minimum of 7-8 calibration standards were used with an r2 coefficient of 0.995 or better and a calibration verification standard (CVS) was analyzed every 15 samples to ensure the integrity of the initial calibration.

Table 4-6: Fit summaries for quadratic model describing MHETase/PETase synergy.

Source	Sequential p-value	Lack of Fit p-value	Adjusted R^2	Predicted R^2
Linear	< 0.0001	0.0013	0.6128	0.5401
2FI	0.1219	0.0017	0.6368	0.5330
Quadratic	< 0.0001	0.6057	0.8383	0.7627
Cubic	0.3880	0.9862	0.8382	0.7433

Significant terms in the quadratic equation:
Intercept, A (PETase load), B (MHETase load), AB, A^2, B^2
Final equation in terms of coded factors (coefficient estimates):
$\ln(\text{mass loss}) = -1.70 + 1.17*A + 0.0226*B + 0.3387*AB - 1.04*A^2 - 0.3562*B^2$
Final equation in terms of actual factors:
$\ln(\text{mass loss}) = -3.94 + 15.54352*A + 1.05674*B + 4.81714*AB - 29.66140*A^2 - 2.53290*B^2$

The coefficient estimate represents the expected change in response per unit change in factor value when all remaining factors are held constant. The intercept in an orthogonal design is the overall average response of all the runs. The coefficients are adjustments around that average based on the factor setting. Final equation based on coded factors can be used to assess the relative impact of each factor, but the equation based on actual factors shouldn't be used to determine the relative impact of each factor.

Samples were initially run undiluted, but this leads to the detection of a very significant doublet signal arising from TPA, and due to a charge species of TPA and matrix effect within the column. Dilution of 5x in dH_2O reforms the doublet peak into a single peak at the RT for TPA. Summing the area under the two peaks gives the same area count as the single peak when considering the dilution factor (within 2%).

Table 4-7: ANOVA for the quadratic model describing MHETase/PETase synergy.

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	33.21	5	6.64	27.96	< 0.0001	significant
A-PETase load	24.54	1	24.54	103.30	< 0.0001	
B-MHETase load	0.0092	1	0.0092	0.0387	0.8459	
AB	1.38	1	1.38	5.80	0.0253	
A^2	6.52	1	6.52	27.47	< 0.0001	
B^2	0.7612	1	0.7612	3.20	0.0879	
Residual	4.99	21	0.2375			
Lack of Fit	0.4732	3	0.1577	0.6289	0.6057	not significant
Pure Error	4.51	18	0.2508			
Cor Total	38.20	26				

Bioinformatics

Sequence selection and conservation analysis 6,671 tannase family sequences were retrieved by a PSI-BLAST search against the NCBI non-redundant database with *Is*MHETase (A0A0K8P8E7.1) as initial query sequence. A total of 3 iterations of the PSI-BLAST search were carried out, and all 6671 hits had E-values of 1e-50 or better. A multiple sequence alignment of the 6,671 tannase family sequences was carried out with MAFFT (12). The amino acid conservation at each site of the multiple sequence alignment was evaluated by computing the relative entropy according to the following equation [146]

$$R.E = \sum_{i=1}^{20} \left(p_i \log \frac{p_i}{p_i^{MSA}} \right)$$

where p_i is the frequency of the i^{th} amino acid in the given site and p_i^{MSA} is the overall frequency of the ith amino acid in the multiple sequence alignment.

Table 4-8: MHETase incubation with MHEI.

250uM Substrate					
Condition		MHETase:MHEI	No Enzyme:MHEI	MHETase:MHET	No Enzyme:MHET
Elapsed Time		24h	24h	24h	24h
Concentration in uM	MHET	-	-	0	283.71 +/- 0.79
	MHEI	261.26 +/- 10.45	250.38 +/- 0.21	-	-
	Terephthalate	-	-	238.04 +/- 2.35	0
	Isophthalate	0	0	-	-
50uM Substrate					
Condition		MHETase:MHEI	No Enzyme:MHEI	MHETase:MHET	No Enzyme:MHET
Elapsed Time		24h	24h	24h	24h
Concentration in uM	MHET	-	-	0	45.17 +/- 0.03
	MHEI	48.70 +/- 0.72	49.00 +/- 0.62	-	-
	Terephthalate	-	-	41.06 +/- 0.04	0
	Isophthalate	0	0	-	-
25uM Substrate					
Condition		MHETase:MHEI	No Enzyme:MHEI	MHETase:MHET	No Enzyme:MHET
Elapsed Time		24h	24h	24h	24h
Concentration in uM	MHET	-	-	0	20.76 +/- 0.07
	MHEI	22.31 +/- 0.43	22.22 +/- 0.003	-	-
	Terephthalate	-	-	15.90 +/- 0.03	0
	Isophthalate	0	0	-	-

Phylogenetic analysis Through keyword searches of the NCBI protein database (https://www.ncbi.nlm.nih.gov/protein), and using BioPython [147], functional annotation for the 6671 sequences was searched. From the sequence description in the NCBI database, 338 and 51 sequences of the 6,671 sequences were clearly annotated as ferulic acid esterases, or as tannases, respectively. Profile hidden Markov models (HMMs) were constructed for ferulic acid esterases and tannases with the dataset of 338 and 51 sequences, respectively, using the HMMER software (version 3.1b2) [148]. Sequence identity thresholds of 95% and 60% were respectively applied to the set of 338 ferulic acid esterases and 51 tannases resulting in a set of 120 sequences (86 FAEs, 31 tannases, *Is*MHETase, and 2 *Is*MHETasea close homologs). The 120 sequences were aligned with MAFFT [149] and phylogenetic analysis with 1000 bootstrap replicates was conducted in MEGA7 [150]. For the phylogenetic tree, the evolutionary distances were computed using the JTT matrix-based method[151].The minimum evolution tree was searched using the Close-Neighbor-Interchange (CNI) algorithm[152] at a search level of 1, and the Neighbor-joining algorithm[153] was used to generate the initial tree. Gaps in the alignment were handled using pairwise deletion. There were a total of 1440 positions in the final dataset.

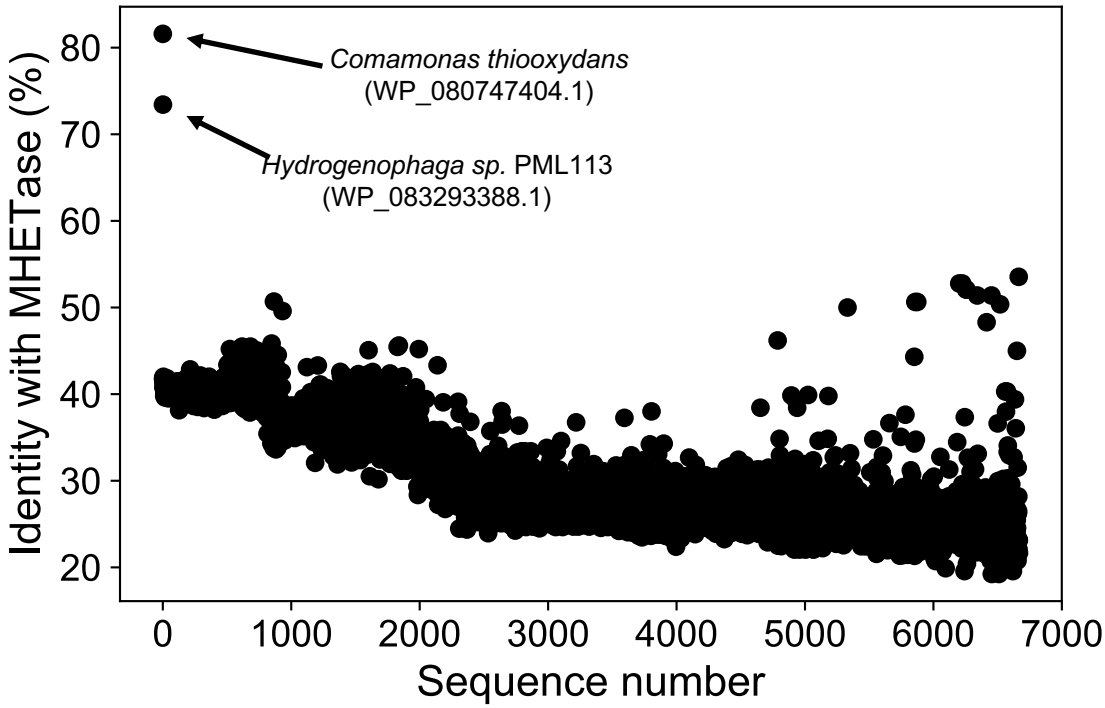


Figure 4-7:. Homology of 6671 tannase family sequences compared to *IsMHETase*. Sequence identities of 6,671 tannase family sequences compared with *IsMHETase* (A0K8P8E7.1). *Comamonas thioxydans* (WP_09747404.1) and *Hydrogenaphaga sp*. (WP_083293388.1) are distinctively the closest homologs to *IsMHETase* in the tannase family.

Molecular docking

MHETase molecular dynamics simulation MHETase structure selected as the initial model for subsequent molecular dynamics simulations. The MHETase structure model was assembled *in silico* using the Schrodinger program's Protein Preparation Wizard in Schrodinger [154–156]. PropKa was used to optimize hydrogen bonds at pH 7.0; OPLS_2005 force field[157] was used to conduct a restrained minimization on all heavy atoms (to ensure less than 0.30 Angstrom deviation from starting structure position).

Ligand structure preparation MHET and MHEI structures were built in Schrodinger using Maestro Workspace tools. All ligand conformations were energy minimized using Schrodinger LigPrep according to OPLS_2005 force field [158]. Ionization states of MHET and MHEI were requested with Epik at pH of 7.0[159] although no additional ionization states were generated.

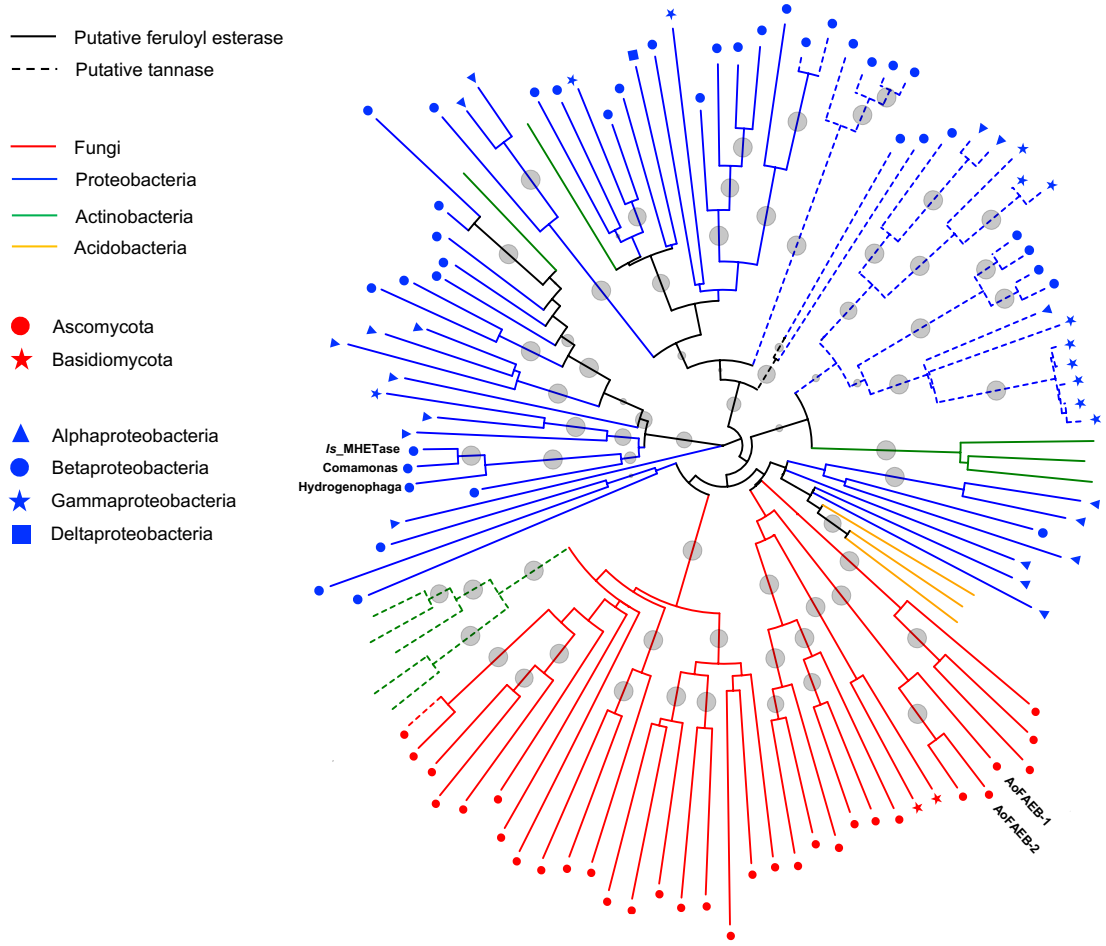


Figure 4-8:. Phylogenetic analysis of 120 tannase family sequences with minimum evolution method and 1000 bootstrap replicates. Nodes with bootstrap values between 75% and 100% are indicated with gray circles with sizes that are proportional to the bootstrap values. Multiple sequence alignment was conducted with MAFFT and the phylogenetic analysis was conducted with MEGA7. *Comamonas* and *Hydrogenaphaga* are the close MHETase homolog sequences, with accession codes WP_080747404.1 and WP_083293388.1, respectively.

Figure 4-8 (continued):AoFAE-B1 and AoFAE-B2 correspond to the *Aspergillus oryzae* ferulic acid esterases, Q2UP89.1 (PDB 3WMT) and Q2UMX6.1 (PDB 6G21), respectively, which in addition to the recently deposited structures, *IsMHETase* (PDB 6QG9) and Fusarium oxysporum (PDB 6FAT), are currently the only sequences with solved crystal structures in the tannase family.

Flexible ligand/flexible receptor docking Induced Fit Docking (IFD) is Schrodinger's flexible ligand/flexible receptor docking tool [160–162]IFD utilizes two other Schrodinger modules, Prime for amino-acid side chain prediction and refinement, and Glide for ligand docking, to achieve binding site flexibility during docking simulations. Ligands were docked into MHETase active site (as established experimentally via co-crystalization with benzoic acid ligand) without trimming of active site residues before initial placement, i.e., only amino acid refinement was done in the active site to model flexibility. This was necessary as attempts to mutate active site residues to alanine and then back-substitute after initial docking (as is often procedure in IFD) would result in a chemically incompetent series of catalytic triad residues. After docking and amino-acid refinement, binding modes were scored and ranked using the Glide XP scoring function. Resulting predicted binding locations were then analyzed to determine if each pose would result in cleavage of an ester bond, such poses were determined to be chemically relevant. Those chemically relevant poses with the lowest predicted binding free energies (i.e. lowest Glide XP score) are discussed in detail in the Results section of Chapter Four.

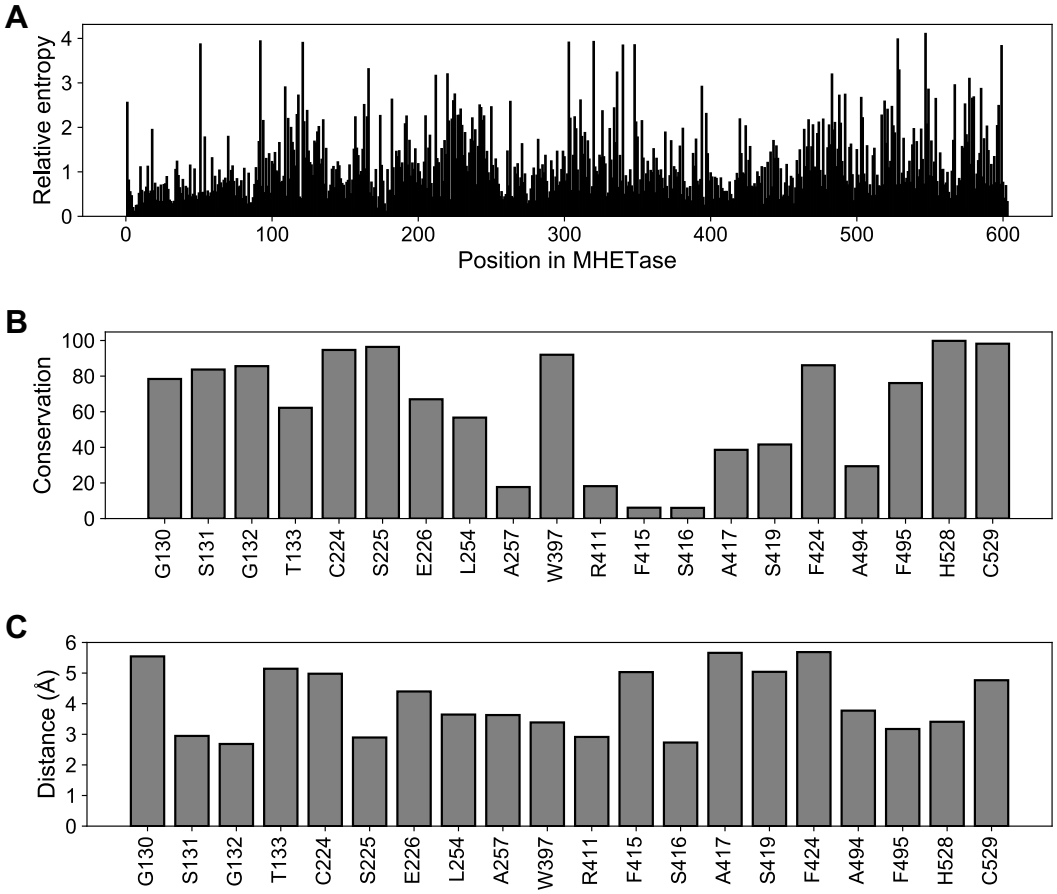


Figure 4-9: Conservation analysis of 6,671 tannase family sequences. (A) Conservation scores (relative entropy) of positions in tannase family sequences plotted against the 603 positions in *Is*MHETase. A higher relative entropy implies a greater level of amino acid conservation in the site. (B) Conservation scores of active-site residues in *Is*MHETase within 6Å of the MHET substrate. Conservation scores are shown as percentiles. R411, F415, and S416 are the least conserved active-site positions and are more variable than 81% of all positions in *Is*MHETase. (C) Closest distance between atoms of *Is*MHETase active-site residues and the MHET substrate.

Molecular simulations

The starting point for molecular dynamics (MD) simulations is chain A of the MHETase structure solved at 1.8 Å resolution that lacks benzoate (PDB code 6QZX). This structure

was chosen because it provides structural information and resolution for the widest range of residues.

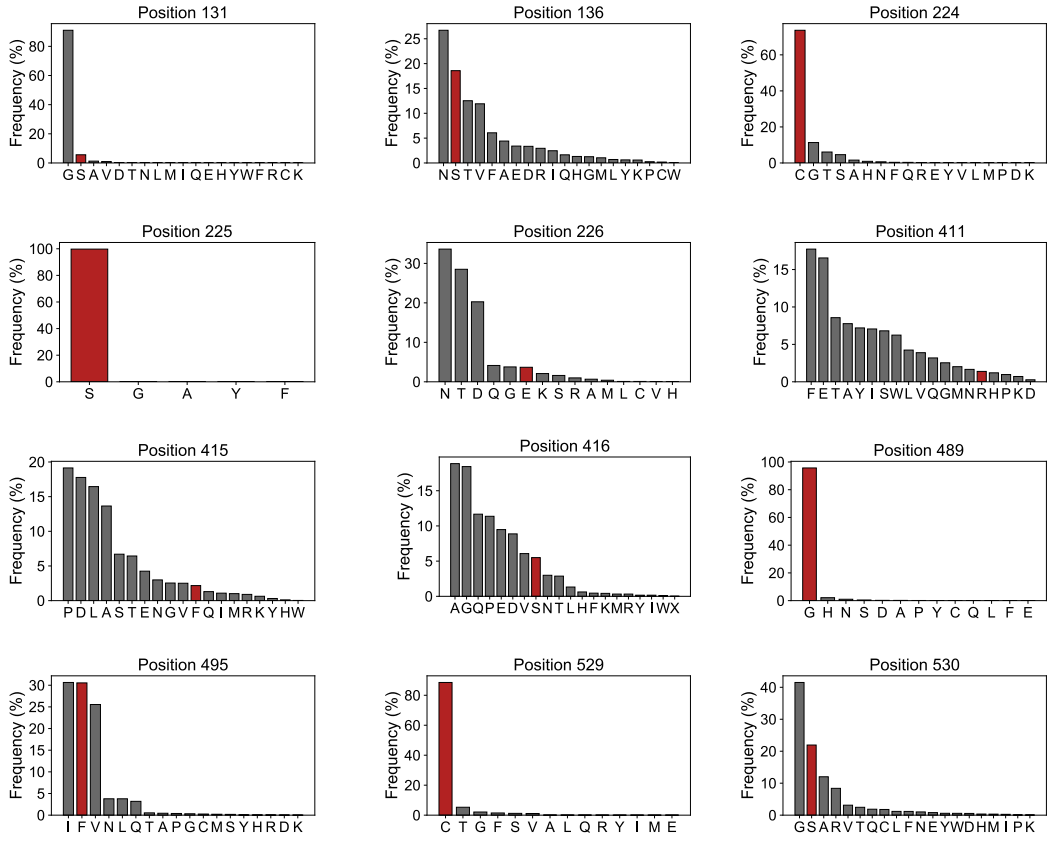


Figure 4-10: Amino acid frequencies of some important positions in MHETase. The frequency of amino acids for each site was determined from the MAFFT multiple sequence alignment of 6,671 tannase family sequences. The positions are named using *IsMHETase* numbering, and the red bars indicate the amino acids in *IsMHETase*.

The bound calcium ion and the crystal waters are maintained (sulfate is deleted). For a variety of residues with alternate conformations, conformation A was chosen for the following: Ser143, Ile149, Ser240, and Asn403. Conformation B was chosen for Ser401and Leu486 (allows for a more relaxed conformation). Initial proposal for protonation states was given by H++ server (pH 7.0, consistent with Yoshida et al.

reaction conditions[1]). Of the acidic residues, Glu230 was determined to be protonated. For histidine residues, His91 and His528 are singly protonated at ND1, His293, His 467 and His488 are singly protonated at NE2, and His166 and His241 are doubly protonated. The overall charge on MHETase with these protonation states is -6; 6 sodium ions were added to the solution phase to neutralize. Five disulfides are formed: Cys51 - Cys92, Cys224 - Cys529, Cys303 - Cys320, Cys340 - Cys348, and Cys577 - Cys599.

All simulations were built using CHARMM version 43a1 and simulated with the CHARMM36 force field for the protein and solvated with TIP3P water molecules. Forcefield parameters for MHET were determined via CGenFF. The simulation box is cubic, with each box side approximately 110 Å long.

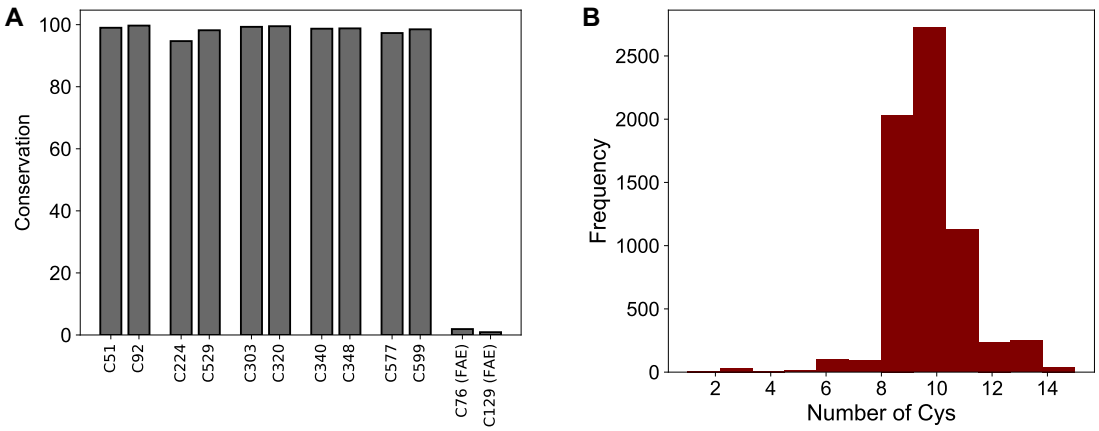


Figure 4-11: Disulfide-bond cysteines in 6,671 tannase family sequences. (A) Conservation of Cys positions forming 5 disulfide bonds in *Is*MHETase. Conservation scores are shown as percentiles. AoFAEB-1 has a 6th disulfide bond between Cys76 and Cys129 which are very variable positions and are less conserved than 98% of positions in multiple sequence alignment. (B) Histogram of Cys occurrence in tannase family sequences showing the rarity of a 6th disulfide bond. Assuming, all Cys form disulfide bonds, less than 8% of tannase family sequences have 6 disulfide bonds.

Approximately 132,000 atoms are modeled in each system. Classical MD simulations of 100 ns in length were run in triplicate for the following three cases: 1) apo MHETase in

which no substrate is bound and no calcium ion is included; 2) apo MHETase (lacking substrate) with calcium ion bound; and 3) MHETase with bound MHET substrate at the active site. For the simulations of T inc complex with MHET, the initial state was prepared as follows: The acidic moiety of MHET is deprotonated. The initial configuration for MHET bound at the active site of MHETase was prepared by aligning the catalytic residues of MHETase with those of PETase bound with a PET dimer at the active site. The atomic configuration of the PET dimer bound in PETase was prepared in a transition path sampling study of the catalytic mechanism of PETase that in turn was prepared from a structure of a PET tetramer bound to PETase from a prior molecular docking study (31). Trimming the PET dimer back to the heavy atoms it shares in common with MHET (maintaining the ester bond closest to the catalytic residues) gave the starting point for MD simulations with MHET bound.

All classical MD simulations were performed at 303K to match the conditions for hydrolytic assays performed by Yoshida et al. (32). Systems were density equilibrated for 1 ns at a constant pressure of 1 atmosphere and constant temperature of 303 K (controlled via the Nosé-Hoover barostat and thermostat); and subsequent production runs were performed with constant volume and temperature (303 K) in NAMD 2.9. All bonded hydrogen distances were constrained utilizing the SHAKE algorithm. The timestep was 2 fs. A nonbonded cutoff distance of 10 Å was utilized, with a switching distance of 9 Å, and a nonbonded pair list distance of 13 Å. The long-range electrostatics were described via the Particle Mesh Ewald (PME) method with a sixth order b-spline, a Gaussian

distribution with a width of 0.312 Å, and 1 Å grid spacing. A velocity Verlet multiple time stepping integration scheme was used, with the full nonbonded interactions

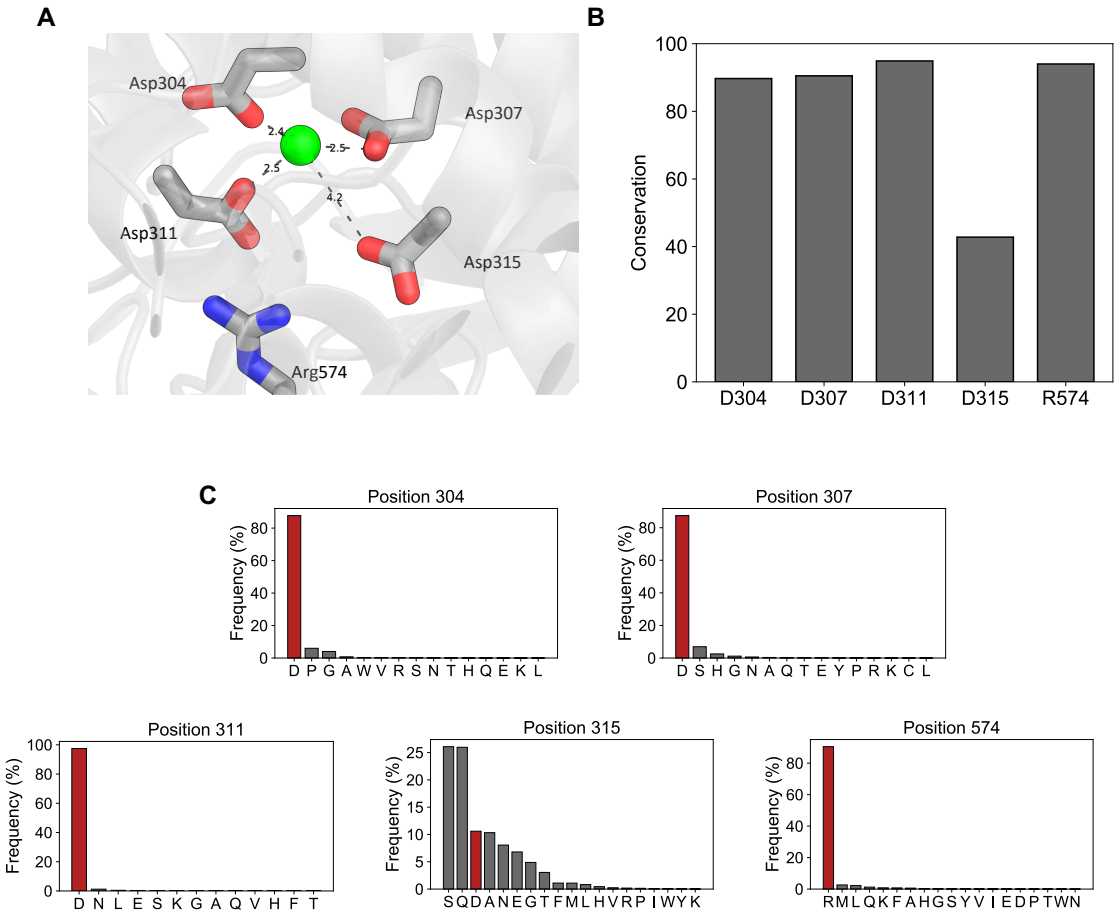


Figure 4-12: Bioinformatic analyses of calcium-binding site residues in *IsMHETase*. (A) Structure of *IsMHETase* at calcium-binding showing the coordinating residues. (B) Conservation of residues coordinating the calcium ion. Conservation scores are shown as percentiles and were calculated from a multiple sequence alignment of 6,671 tannase family sequences. With the exception of Asp315, which is less conserved than 58% of all position in *IsMHETase*, all calcium-coordinating residues are strongly conserved (more conserved than 89% of positions in *IsMHETase*). (C) Amino acid frequencies at calcium-coordinating position in multiple sequence alignment. About 90% of tannase family sequences have an amino acid other than Asp at position 315.

evaluated every 2 timesteps, full electrostatics interactions evaluated every 4 timesteps, and 20 timesteps between atom reassignments.

Hybrid quantum mechanics/molecular mechanics (QM/MM) simulations were performed in Amber 12. Following classical simulations, the CHAMBER utility of ParmEd version 3.0.3 was used to convert the CHARMM coordinate, topology, parameter, and protein structure files to AMBER formatted coordinate and topology files for the Sander program of AMBER version 13.11-12 The AMBER software was used to carry out all QM/MM calculations with the Self-Consistent Charge Density-Functional Tight-Binding (SCC-DFTB) semiempirical QM method using the Third-Order Parameterization for Organic and Biological Systems (3OB) to describe the QM region.13 An 8 Å cutoff was used for short-range and PME used for long-range interactions. Periodic boundary conditions were used, the timestep reduced to 1 fs, and SHAKE was applied only to hydrogen in the MM region. The Langevin thermostat and barostat were specified the same for the QM/MM simulations as they were for the MM simulations.

The QM region includes the MHET substrate and the three catalytic residues (Ser225, Asp492, and His528, each cut across the Cα/Cβ bond). For step 1 of the catalytic mechanism (acylation), there are 46 atoms in the QM region, with a charge of -2. Hydrogen link atoms are utilized where covalent bonds cross the boundary between the QM and MM regions, as implemented in Amber. For step 2 of the catalytic mechanism (deacylation), the ethylene glycol product is removed (as it was shown to leave the active site after the acylation step) and a single water molecule is added to the QM region, giving 39 atoms in the QM regions still with a charge of -2. The timestep for QM/MM

simulations was 1 fs; hydrogen atoms in the QM region were not constrained by the SHAKE algorithm. Temperature for the QM/MM was 310 K.

CHAPTER FIVE

DISCUSSION AND CONCLUDING REMARKS

The attention PETase has drawn from general public and researchers alike can be attributed to its applicability in industrial settings. WT PETase, however, as Austin *et al.* stated, is not only not optimized for efficient PET degradation at its working temperature of 30ºC but is also inactive at the T_g of PET (70-80°C). Therefore, engineering of a thermostable PETase with higher PET degradation efficiency is the ultimate goal for industrial applications and, of course, for embedding novel methods of recycling plastics in circular economies[163]. While PETase is not yet fully ready for its debut in an industrial setting, a number of promising structural and enzyme engineering studies have been submitted [2,66–69]. Kawai et al. expressed concern for the ability of PETase to degrade lcPET at 30°C[73]. Here, we have demonstrated that PETase is not only capable of such process, but can also hydrolyze, to a small extent, boPET films, as shown in our nano-scale studies and product release observed in HPLC. Data presented in Chapter Three provided further insights into PETase's ability to hydrolyze different PET substrates, novel and nano-scale approaches to assess the morphological features of PET relevant to its degradation by PETase, and enzyme kinetics parameters when using soluble substrates such as *p*-NP esters and BHET, one of the byproducts of PET hydrolysis.

We have attempted to answer previous concerns raised by Kawai et al. regarding PETase's ability to hydrolyze lcPET. We found that WT PETase and its W159H/S238F double mutant are successful at deconstructing lcPET. Furthermore, PETase, similar to

other PET degrading cutinases, exhibits a preferential hydrolysis process on different sides of lcPET coupons. Preferential hydrolysis is known to be a factor in many other polyester degrading enzymes, including *Fusarium solani* cutinase, *Rhizopus oryzae* lipase, *Thermobifida cellulosytica* cutinase and others[125,164]. This preferential hydrolysis depends on the topology and the crystalline morphology of the PET substrate. Although it is possible to inspect PET morphology using atomic force microscopy (AFM), this approach does not provide a direct observation of the lamellae structures of the crystalline regions[93,165]. We have shown here that lcPET thin films exhibit an increase in lamellar width and polymer chain order following incubation with PETase. Conversely, PETase treatment of boPET thin films leads to a decrease in lamellar width, and polymer chain order, suggesting that crystalline regions of boPET can be hydrolyzed by PETase. Results of this study regarding lcPET and boPET at the nano-scale are consistent with Austin *et al.*'s results[2] , which explored the crystallinity reduction in sPET (~14% crystallinity, with an absolute reduction of 1.5%). Furthermore, this study provided a basis for detecting hydrolysis products from boPET degradation using HPLC. We are exploring the idea that during PETase hydrolysis of lcPET, a substrate with low crystallinity (4%), PETase may be encountering amorphous regions more often than crystalline regions, leading to a greater overall hydrolysis of amorphous regions compared to their crystalline counterparts. Therefore, for lcPET, an increase in the percent polymer chain order and crystallinity is observed. In the case of boPET, hydrolysis of less ordered regions is initiated by PETase, which presumably provides access to the less ordered, more hydrolytically susceptible regions of the lamellae, and

thus leading to a decrease in polymer chain order. Thus, while enzymatic hydrolysis of boPET can occur to some extent, it is however necessary to reach regions of lower order within the lamellae regions to initiate hydrolysis of the boPET polymer crystalline regions, which leads back to the idea of increasing the reaction temperature, preferably closer to the T_g of the polymer to maximize hydrolysis efficient. The approach presented in Chapter 3 which include examination of PET films of various crystallinities and morphologies pre and post digestion with RuO_4 staining, is a novel method that has not been fully explored in the 20 years of research on esterase/cutinase degradation processes, to the best of our knowledge. Nano-scale approaches to enzymatic PET degradation were attempted previously, using quartz crystal microbalance with dissipation monitoring (QCM-D)[166], but this method does not provide any visual data on what might be occurring at the nano-scale. While our method still needs further optimization for substrates like PEF, we believe that, in the future, the nanoscale approach to enzymatic PET degradation will provide crucial information about the molecular processes underlying hydrolysis of aromatic polyesters.

Visualizing PETase activity at the chemical resolution level, i.e. using substrate and enzyme labeling (such as GFP labeling) techniques also provides a promising avenue, especially with respect to better understanding the co-localization of PETase and MHETase for synergistic interactions between the two enzymes. While no results regarding chemical resolution studies were shown here, efforts towards enzyme labeling are being carried out with the help of Dr. Lahiru Jayakody and Dr. Bryon Donohoe at NREL, with Ms. Rita Clare currently continuing this endeavor at NREL.

On a micro-scale, sPET degradation by PETase generates similar patterns to those observed by Austin et al. Based on these hydrolysis patterns and the formation of "islands", we suggest that the latter represent spherulite structures present in sPET. Spherulite structures of PET have been observed with SEM, as reported by Baldenegro-Perez [167]. SEM micrographs of polybutylene succinate (PBS, Table 1-1) pre and post-digestion by PBS degrading enzymes also shows clear formation of spherulites (data not shown), which generally cannot be fully degraded by PETase and are also left mostly undigested (Figure 3-5). Eventually, as shown by DSC, hydrolysis of crystalline regions occurs by taking advantage of access to less ordered regions in the lamellae. Efforts to further elucidate the morphology of sPET at micro and nano scale are ongoing.

Prediction of the thermal stability of enzymes is a tricky endeavor, as the direct relationship between protein structure and thermal stability remains to be fully understood. It is, however, quite easy to envision that, once an algorithm predicting thermal stability is developed, optimization of biotechnological processes, such as improved enzymatic PET degradation, would be an relatively effortless venture. While such algorithms exist, there is still room for improvements, as other factors such as pH, ionic content of the buffer, and even protein concentration should be taken into account as those also impact protein thermal stability[168] . In spite of these limitations, the SCooP software has been shown to predict the thermal stability of WT PETase (6EQE) quite well. The double point mutations introduced to WT PETase which have yielded the W159H/S238F PETase variant, have resulted in a more efficient PETase enzyme and an improved PET hydrolysis at 40°C. While the mass loss at 40°C is not significantly

different from the mass loss at 35°C for W159H/S238F, this mass loss is higher than that observed for the WT PETase reaction at 40°C, a temperature where WT enzyme inactivation becomes significant. As mentioned in Chapter Three, the effect of atomic packing might be responsible in part for the increased thermostability of W159H/S238F observed at elevated temperatures, as a recent study has shown that increased atomic packing positively influence the structural stability of PETase [130]. Furthermore, a smaller active site cleft directly mimicking that of a protein of thermophilic origin (*Tf*Cut) strengthens this rationale.

Hydrolysis of BHET by PETase possibly interferes with efficient PET degradation since BHET is a byproduct of PET hydrolysis. WT PETase appears to have a slightly higher affinity for BHET than the W159H/S238F PETase variant, which may provide another rational for the improved hydrolysis of PET observed with W159H/S238F PETase. A very interesting idea to follow up would be to employ protein engineering to design an active site environment similar to that of MHETase, while maintaining (or, better yet, improving) the PET degradation process by minimizing the hydrolysis of BHET by PETase. Furthermore, designing a protein that can hydrolyze BHET (or only BHET) and examining its possible synergistic effects on PET degradation with PETase and MHETase would broaden our ability to maximize enzyme synergistic interaction for a more efficient PET degradation process.

While the ability of PETase to hydrolyze aromatic polyesters was not investigated extensively, the concept of substrate-specific enzymes for stepwise purification of PET degradation monomers in an industrial setting is an intriguing idea. The three putative

PETases examined here did not show any activity towards PET, PBS and PLA, indicating that they in all likelihood do not possess the same capabilities as from *I. sakaiensis* PETase, which is able to hydrolyze such polyester substrates. However, it is also possible that the proteins have different working temperatures than the PETase from *I. sakaiensis,* although all the putative PETases were examined for activities at both 30°C and 40°C. A protein with high sequence similarity to PETase, which can degrade PBS but not PLA, may be an invaluable tool for assessing the intricacies of aromatic-only polyester hydrolysis by PETase classes of enzymes.

Chapter Three shed further light on the mechanism of enzymatic PET hydrolysis by PETase, and possible routes for surface modification to enhance PETase-mediated PET degradation. On the other side of the story, in Chapter Four, lies a large collaborative effort, which describes in great detail the structure, bioinformatics analysis, biochemical features, and potential avenues to take advantage of MHETase to enhance the process of PET degradation by PETase. While MHETase has not been as intensely studied as PETase, the MHETase enzyme represents in a way "the second half of the puzzle", as full hydrolysis of PET into its constituent monomers had not been observed when MHETase is absent. Bioinformatics analyses, accompanied by structural methods revealed that MHETase is a member of the Tannase family of enzymes, whose members range from tannases to feruloyl esterases, including the *Aspergillus oryzae* feruloyl esterase *Ao*FaeB-2 (PDB ID: 6G21). Furthermore, the high-resolution crystal structure of MHETase reveals that the α/β hydrolase domain of PETase is retained, but with an interesting addition to it, i.e. the presence of a "lid domain" which wraps around and

forms a hydrophobic substrate binding pocket (Figure 5-1). Specifically, the lid domain of MHETase lid domain is inserted, with respect to the location of the loop structures of PETase, within a 7-residue loop region that connects α-helix α_4, and the PETase β_6 β-strand. In contrast, in MHETase crystal structure, this 7-residue loop is extended into a ~200-amino acid-long domain, which then connects back to the 6th β-strand of the α/β hydrolase domain of MHETase (Figure 5-2). Based on the presence of the lid domain in MHETase, we proposed to engineer enzyme variants of PETase with a lid and of MHETase with no lid, and several variants of "lidless" MHETase enzymes were generated as described in Chapters 2 and 4. MHETase additionally possesses 10 cysteine residues, all of which partake in the formation of 5 disulfide bonds spread out through the protein lid and α/β hydrolase domains. Bioinformatics analyses revealed that the disulfide bonds formed by these cysteines are conserved within the Tannase family of enzymes, but several members contain even more, amounting to the presence of six or seven disulfide bonds. Interestingly, one of these disulfides is located in an equivalent position to that of Trp159, which has been replaced by a histidine in the W159H/S238F PETase double mutant and which flanks a catalytic serine residue in PETase. On the subject of the MHETase active site, bioinformatics analyses also revealed conserved amino acids at distal positions from the active site.

Analysis of the structural and bioinformatical features of MHETase resulted in engineering new enzyme variants, with the goal of assessing the impact of adding or deleting disulfide bonds, or removing the lid domain, on MHETase enzyme function. One protein variant in particular, PETlid, informally referred to as a "Frankenstein's

PETase", and consisting of a PETase enzyme comprised of a MHETase lid domain, is promising in a sense that it remains in the soluble fraction when overexpressed in *E. coli,* although the protein yields are low. If PETase could acquire MHETase activity by adding a lid domain while retaining its original PETase activity, such characteristics would be a great demonstration of the modularity of enzyme function for plastic degrading polyesterases. Our efforts to fully characterize these PETase and MHETase mutants are ongoing.

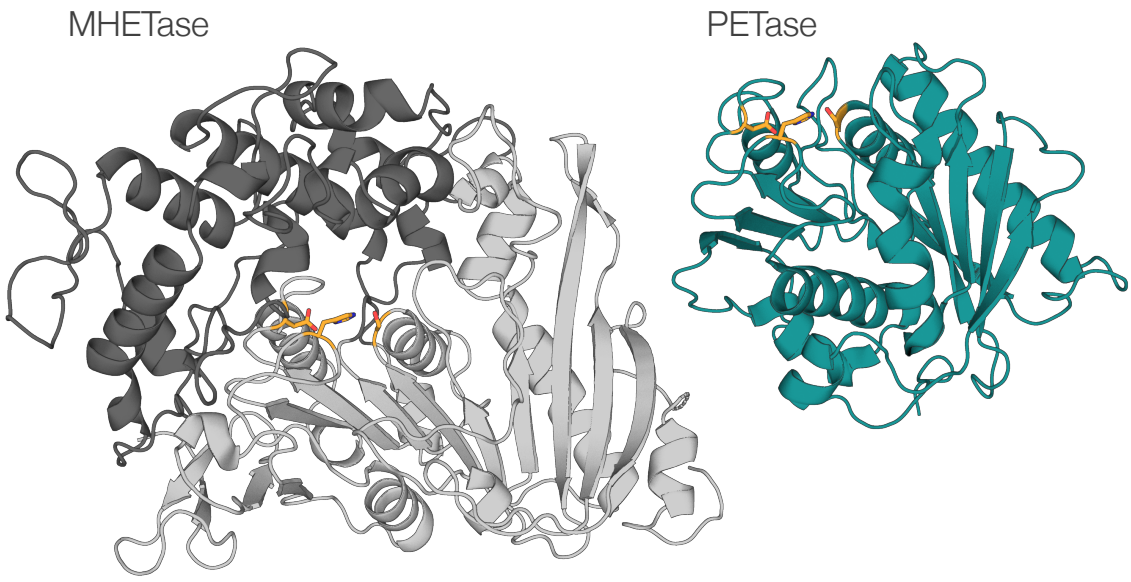


Figure 5-1: Comparison of MHETase with its lid domain (black) and hydrolase domain (light grey) and PETase (teal). Catalytic triads of both enzymes are colored orange.

Concluding remarks

Overall, data presented in this dissertation represent valuable steps towards generating better enzymes for efficient and sustainable chemical recycling of PET. As previously mentioned, while PETase is not ready to be applied at an industrial level,

working towards increasing its activity, and exploration of surface modification methods can pave the way for industries which aim to devote their resources enzymatic plastic recycling.

Although the introductory and informational sections in Chapter One paint a worrisome picture about the current plastics problem, using organisms genetically engineered to degrade PET in wastewater treatment plants, and investments toward chemical recycling can help us build a better, more sustainable future.

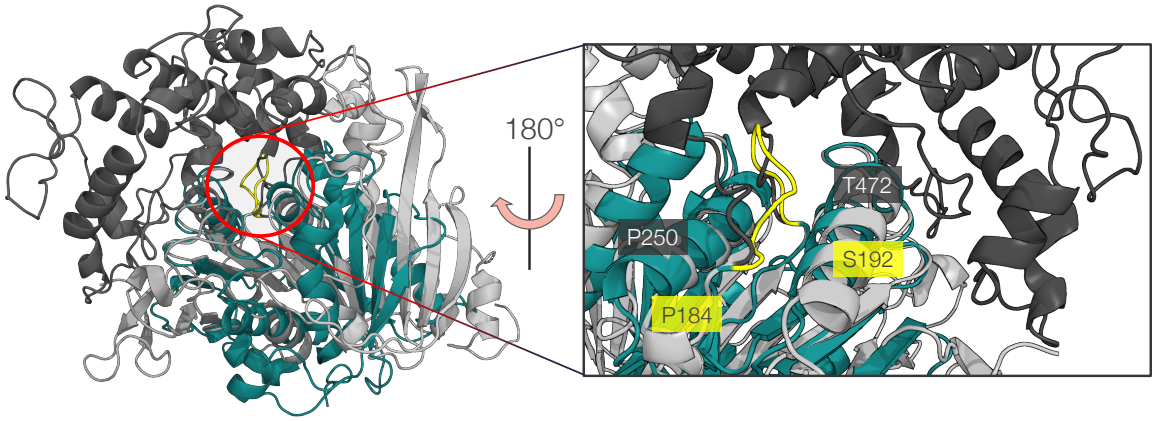


Figure 5-2: Structural alignment of PETase (teal) and MHETase with its lid (black) and α/β hydrolase domain (light grey). Zoomed in area depicts the seven-residue loop (shown in bright yellow) in PETase spanning Pro184-Ser192, which is replaced in MHETase by a lid domain (black), spanning Pro250-Thr472.

APPENDICES

APPENDIX A

DEVELOPMENT OF METHODS FOR 1H NMR METABOLOMICS FOR HUMAN SERUM

This section briefly describes methods of metabolite extraction and metabolite analysis from human plasma as part of a collaborative study between Montana State University (MSU) and University of Texas Southwestern Medical Center (UTSW).

Fast development of -omics based approaches for modeling disease state is gaining increased attention since the appearance of first -omics discipline, genomics[169]. Combinatorial approaches consisting of genomics, proteomics, metabolomics and even microbiomics to model simple infection to complex multifactorial disorders has become a desirable approach for clinical diagnosis[170–175].

In the collaborative project between MSU and UTSW, we have looked extracted serum metabolites from 250 patients, who were a part of a study regarding the outcomes of using a combinatory-treatment approach to change depression outcomes[176] . The ultimate goal of this project was to determine whether metabolites of significance can be used to predict outcome of diseases. While the field of metabolomics (or most -omics in general) cannot yet make those predictions, developing methods of 1H NMR metabolomics for analysis is of great importance. To that end, we extracted, analyzed and quantified the metabolites of 250 patients at t=0, and at t=12 weeks post treatment and patient groups were separated by which treatment they received. We've additionally performed preliminary PCA analyses on the dataset, to identify differences in metabolomes between the two groups. The preliminary analyses performed here cannot make predictions.

Materials and Methods

900 µL of acetone were mixed in with 180 µL serum sample from each patient and vortexed for 30 seconds. Samples were then vacuum dried using a speed-vac. All NMR spectra were collected at 298 K, using a Bruker as previously described[177]. After data collection, individual metabolites were assigned and quantified for each patient on Chenomx. Mean centered and log transformed metabolite data for all 250 patients were analyzed by MetaboAnalyst.

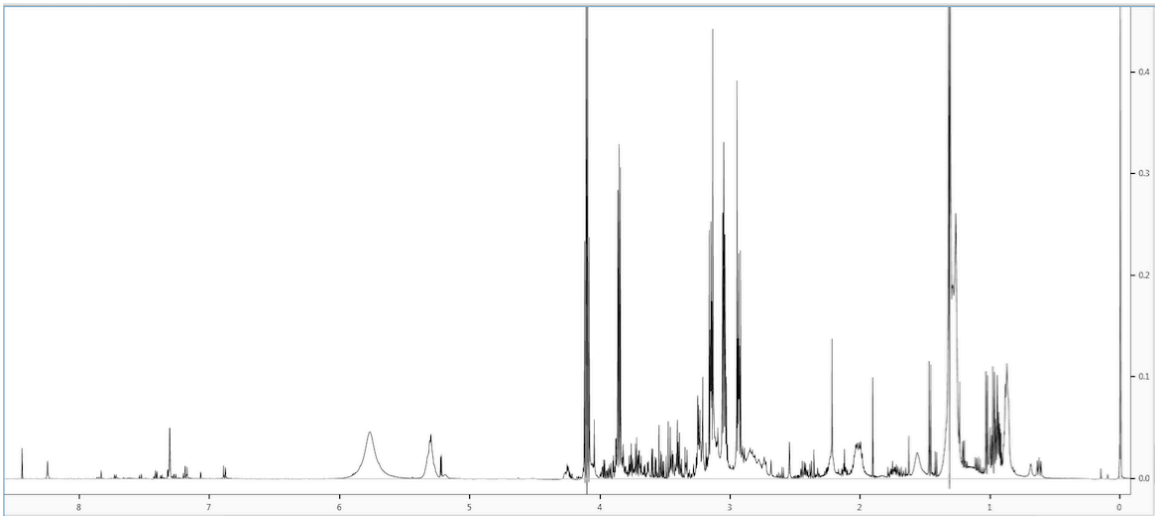


Figure A-1: Representative NMR spectrum from a serum sample.

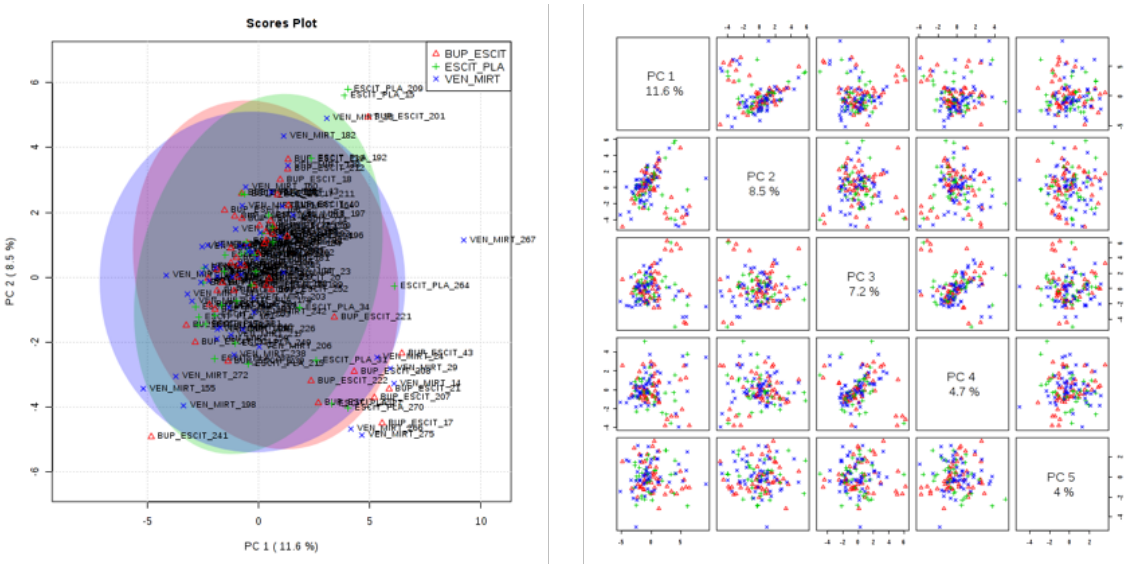


Figure A-2: Left: 2D - PCA plot of the metabotypes of the patients from t=0 w, Right: Variance of the first 5 principal components. n=152 PC1: 11.6%, PC2: 8.5%.

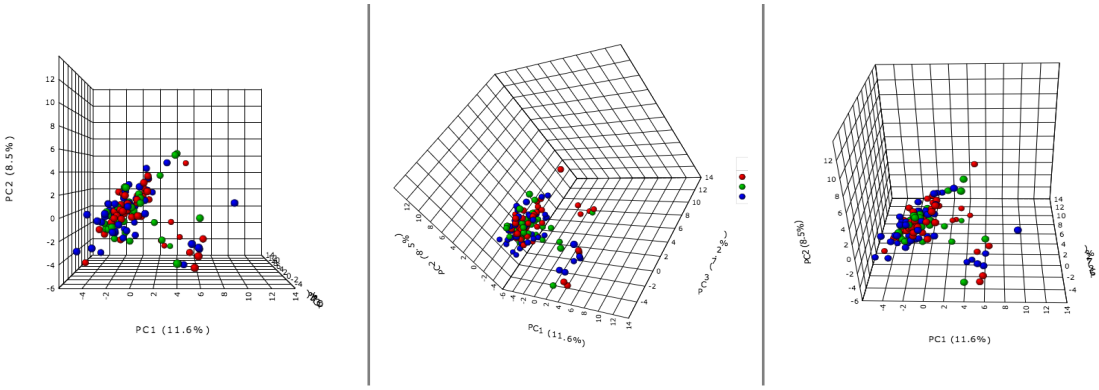


Figure A-3: 3D – PCA plots of the metabotypes of the patients from t=0 w from different angles. n=152; PC1: 11.6, PC2: 8.5, PC3: 7.2%

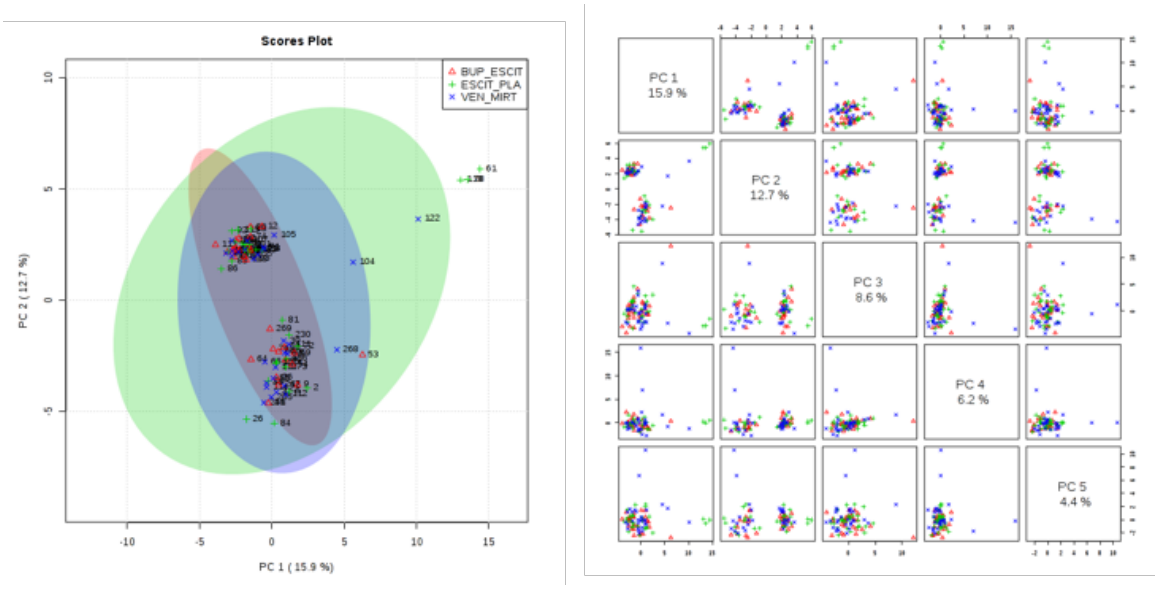


Figure A-4: Left: 2D - PCA plot of the metabotypes of the t=12w patients, Right: Variance of the first 5 principal components. n=88 PC1: 12.7%, PC2: 15.9%.

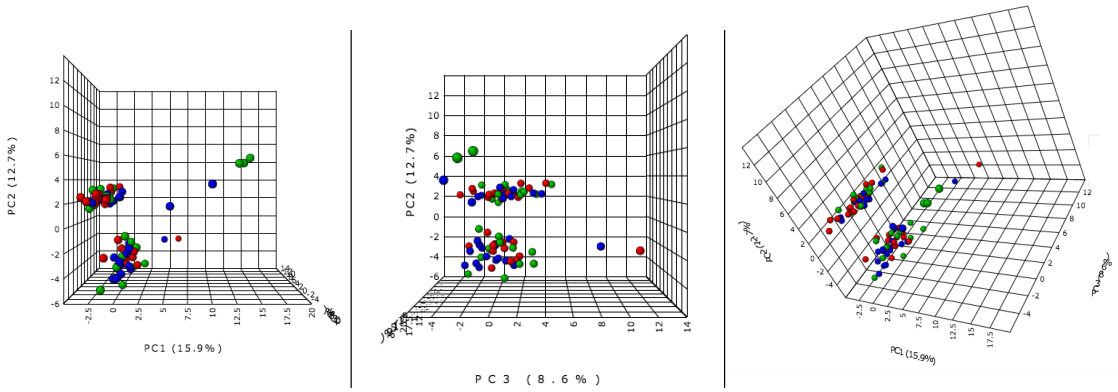


Figure A-5: Figure A-2: 3D – PCA plots of the metabotypes of the patients from t=0 w from different angles. n=88; PC1: 12.7%, PC2: 15.9%, PC3: 8.6%.

Table A-1: Comprehensive list of metabolites identified in Chenomx.

#	A-H	I-X
1,7-Dimethylxanthine	Acetaminophen	Ibuprofen
2-Hydroxybutyrate	Acetate	Isoleucine
2-Hydroxyphenylacetate	Acetylsalicylate	Kynurenine
2-Oxoisocaproate	Alanine	Lactate
3-Hydroxybutyrate	Arginine	Leucine
3-Methyl-2-oxovalerate	Ascorbate	Lysine
4-Hydroxyphenyllactate	Asparagine	Malonate
4-Pyridoxate	Betaine	Melatonin
	Caffeine	Methionine
	Carnitine	N-Acetylaspartate
	Choline	O-Acetylcholine
	Creatine	Ornithine
	Creatinine	Phenylalanine
	Dimethyl sulfone	Proline
	Formate	Pyroglutamate
	Fructose	Pyruvate
	Glucose	Salicylate
	Glutamate	Salicylurate
	Glutamine	Serine
	Glutarate	Succinate
	Glycerol	Taurine
	Glycine	Threonine
	Hippurate	Trimethylamine N-oxide
	Histamine	Tryptophan
	Histidine	Tyrosine
	Homoserine	Urea
	Hypoxanthine	Uridine
		Valine
		Valproate
		Xanthine
		Xanthosine
		sn-Glycero-3-phosphocholine
		pi-Methylhistidine
		tau-Methylhistidine

APPENDIX B

CRENARCHAEAL ADAPTIVE IMMUNE SYSTEM FUNCTION AND REGULATION BY THE PROTEIN, CSA3

Outline of this project

The work presented in Appendix A expands further into the regulatory mechanisms involving CRISPR-*cas* adaptive immunity of Crenarchaea and extend newly-gained knowledge on Archaeal families and possibly to Bacteria. Specifically, a structure-function relationship driven work was conducted for the protein Csa3 in *Sulfolobus islandicus* M16.4. The endgame of this work was to provide a comprehensive mechanism by which *S. islandicus* M16.4 regulates protein expression upon invasion of alien sequences. To this end, structure of the putative transcriptional regulator, Csa3 from *S. islandicus* was solved, its ability to activate (or repress) transcription was investigated, and multiple small molecule binding partners of Csa3 were explored and cocrystallization attempts were made with the aforementioned small molecule binding partners. This work resulted in a publication on RNA Biology[100] with another anticipated research paper in the works, related to the ligand binding properties of Csa3.

Discovery of Prokaryotic Adaptive Immunity

All domains of life on Earth are constantly under the threat of infection by viruses and other pathogens. As a result of this struggle between pathogens and their hosts, hosts have evolved innate and adaptive defense mechanisms. Until about a decade ago, Eukaryotes were thought to be the only domain of life that harbored adaptive immune systems. After the discovery of CRISPR (Clustered regularly interspaced short palindromic repeats) loci in Prokaryotes, several studies showed that CRISPR arrays had

several identical sequences to some foreign genetic elements. It was then hypothesized CRISPR arrays may have been a part of prokaryotic adaptive immune systems [178–180]. In 2007, this hypothesis was demonstrated in *Streptococcus thermophilus*, in which the fragments of invading DNA were inserted into the CRISPR array, resulting in phage-resistance. In addition, it was demonstrated that *cas* genes are absolutely necessary for the acquired phage resistance. We now know that Archaea and Bacteria use CRISPR-*cas* adaptive immune system for protection against viruses and other mobile genetic elements. About 50% of Bacteria and 90% of Archaea encode CRISPR-*cas* loci, according to the CRISPRdb [181].

CRISPR-*cas* Mechanism, Design and Diversity

CRISPR arrays consist of an AT-rich leader sequence, followed by short repeat sequences separated by distinct spacers from invading nucleic acids. Spacers from CRISPR arrays are used as transcripts that recognize invading viruses and degrade them [182]. CRISPR-associated (*cas*) genes are often found in the vicinity of CRISPR loci and Cas proteins are the main protein machinery in mediating the adaptive immune response against invading genetic elements. Cas proteins involved in CRISPR immunity are very diverse. In 2005, Haft *et al.* identified 45 *cas* gene families which include "core" cas genes (Cas1-Cas6), subtype-specific genes (*cse1-4, csy1-4, csn1-2, csd1-2, cst1-2, csh1-2, csa*1-5, csm1-5), Repeat-Associated Mysterious Protein (RAMP) superfamily (*cmr1*-6) and seven gene families that lacked a contextual pattern (*csx1-7)* [183]. In order for CRISPR-*cas* adaptive immunity to defend the host against invading genetic elements,

Cas proteins and host cell machinery have to successfully achieve the following major stages: (1) Acquisition, (2) Biogenesis and (3) Interference [184]. Each step employs various Cas proteins and necessary for the acquired immunity. Presence of multiple CRISPR loci, the number of highly diverse Cas proteins, existence of incomplete, putative and horizontally transferred CRISPR-*cas* loci, make evolutionary classification of CRISPR-*cas* very challenging [185]. Current classification of CRISPR-*cas* systems have evolved over time; and the most current classification of CRISPR-*cas* loci across domains currently have two major classes, five types, and sixteen subtypes [186]Type IV and type V systems contain either putative or incomplete CRISPR-*cas* loci. Types I, II and III can be found in Bacteria; and types I and II are found in Archaea. Archaea lack the RNAse III enzyme that is required for processing double stranded RNA, which is an essential component of Type II systems [185]

Key steps of CRISPR-cas immunity

Acquisition

When an invading nucleic acid is present in the cell, small fragments of invader DNA, called protospacers, are inserted into CRISPR array in a leader-proximal fashion [178]. Protospacers are flanked (upstream or downstream) by 2-5 nt long recognition sequences called Protospacer-adjacent motif (PAM). PAM appears to be responsible for not only the recognition of protospacers, but also the orientation of the spacer as it is inserted into the CRISPR-array [187]Cas1 and Cas2 were identified as the two proteins that were essential for acquisition, but are not required for biogenesis and interference stages.

It is also worth noting that Cas1 and Cas2 are present in virtually all CRISPR-*cas* systems. Cas1 and Cas2 are commonly found encoded within the same operon along with a few other subtype-specific *cas* genes.

Biogenesis.

Expression and maturation of crRNAs occur in two steps. First, a long, primary transcript of the CRISPR loci, called precursor crRNA (pre-crRNA) is produced from the promoter that is within the leader sequence. Second, pre-crRNA is cleaved into mature crRNA.

Interference.

Recognition of invading genetic elements and their degradation is the final step in CRISPR immunity. This step includes scanning protospacer sequences that are complementary to crRNA, and the degradation of the target by specific Cas nucleases [188]. Csm and Cmr complexes of type III systems appear to degrade both DNA and RNA. Csm complexes from *Staphylococcus epidermidis* III-A system have recently been shown to independently degrade both dsDNA and ssRNA *in vitro* [189]. In the Csm complex, Cas10 is responsible for independent DNA cleavage, whereas Csm3 is responsible for ssRNA cleavage into 6 nt fragments. In type III-B (Cmr) systems, Cmr1 and Cmr6 recognizes the target DNA, whereas Cmr4 mediates the target RNA cleavage in 6 nt intervals [190].

Archaeal Adaptive immunity

Archaea harbor type I and type III systems, and type III systems are commonly found in thermophiles [191]There's a strong bias towards types I-A, I-B and I-D among Archaea; there are very few I-C and I-E, and there is no I-F subtype.

CRISPR-Cas Immune Systems of the Order *Sulfolobales*.

The order *Sulfolobales* has important model organisms for the development of Archaeal physiology, as well as CRISPR-Cas[180,187,192–194]. Most of the *Sulfolobus* species often carry multiple CRISPR loci and multiple copies of type I and type III proteins. *S. solfataricus* P2 has 6 CRISPR loci, and multiple copies of type I-A and III-B systems. *S. islandicus* M16.4 has 2 CRISPR loci and type I-A, and type I-B systems.

Regulatory proteins of CRISPR-*cas* in *Sulfolobales*.

In 2005, Haft *et al.* identified several putative CRISPR-*cas* regulators both from Archaea and Bacteria. In 2014, Vestergaard *et al.* have identified l35 putative regulatory proteins from Archaea. It was also identified that several Csa3 proteins were encoded in the vicinity of adaptation and interference operons and gene cassettes [191]. Csa3 was first identified by Haft *et al.* (2005) and belongs to COG1517, in which the structures of 4 proteins (Csm6, Csx1, VC1899 and Csa3) with distinct C-terminal domains have been determined.

Regulatory Protein Csa3

Structure of Csa3 from *Sulfolobus solfataricus* (denoted *ss*Csa3) reveals a two-domain, dimeric protein [99]. C-terminal domain of [195]ssCsa3 exhibits a common DNA binding domain. N-terminal domain of Csa3 shows a putative ligand-binding domain that is now named as "CRISPR-*cas* associated Rossmann fold" (CARF) [196].

C-terminal domain of Csa3 is a MarR-like wHTH DNA-binding domain. wHTH domain is part of the helix-turn-helix superfamily of proteins that are involved in DNA recognition and binding [195]. wHTH domain of *ss*Csa3 consists of 4 alpha helices and two beta strands. Helices αC1, αC2 and αC3 form a right-handed helix bundle. The bundle is followed by a β-hairpin and another α-helix that form the "wing" of the C-terminal domain. wHTH domain of Csa3 makes it an excellent candidate for regulating CRISPR-cas. Csa3's role in regulating expression of adaptation genes has been shown in *Sulfolobus islandicus* REY15A (denoted Csa3a) [197]. Csa3a was able to bind promoter sequences of *csa1* and *cas1*. In addition to that, overexpression of Csa3a led to increased transcription of the adaptation genes and caused an uptake of *de novo* spacer sequences mostly from the overexpression vector, as well as the inserted spacer sequences from the host CRISPR array. However, the mechanism by which Csa3a received a signal to activate protein expression, and its potential effects on regulating other stages of CRISPR-*cas* immunity remain unknown.

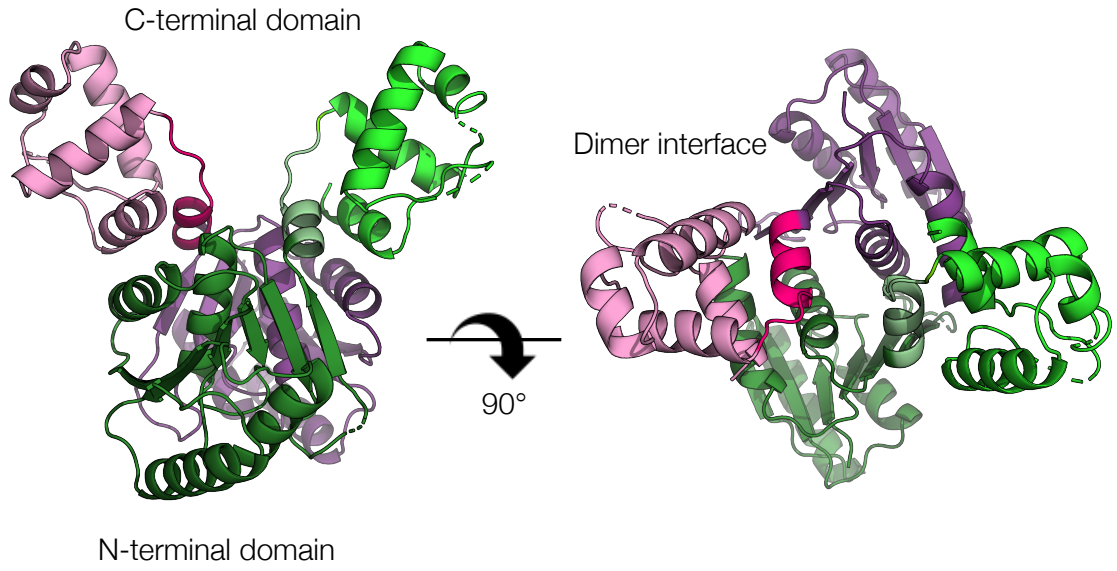


Figure B-1**:** *si*Csa3 dimer. N-terminal domain, linker and C-terminal domain are labeled forest green, light green and TV-green in chain A and deep purple, hot pink and light pink in chain B, respectively. "Top down" view of the dimer is shown to affirm the dimer interface.

To that end, we expressed, purified and crystallized Csa3 from *Sulfolobus islandicus* and solved its crystal structure via molecular replacement. N-terminal domain of siCsa3 shows a putative ligand-binding domain that is now named as "CRISPR-*cas* associated Rossmann fold" (CARF). The Rossmann fold consists of three consecutive β-strands that run parallel to each other, as a result of the right-handed connections made by α-helices [198](Figure B-2A). Rossmann folds are typically seen in dinucleotide binding domains, in which two Rossmann folds form a mixed β-sheet with α-helices on both sides of the sheet (Figure 3B). In many ways, CARF domains resemble dinucleotide-binding domains with two Rossmann folds, but there are key structural differences. N-terminal CARF domain of Csa3 is a 6-stranded, mixed β-sheet in which the first five β-strands of the domain run parallel to each other, as a result of right-handed connections made with

α-helices. The last two strands, βN5 and βN6, make a reverse turn and βN6 runs antiparallel to the rest of the strands in the sheet. As a result of this reverse turn, a right-handed α-helix is lost. This is the defining feature of the difference between CARF domain of Csa3 and dinucleotide binding domains (Figure B-2B). However, when Csa3 dimer is formed, the αN4' of the symmetry related subunit runs parallel to the αN4, covering βN5 and βN6 and restoring the structural element of the dinucleotide binding domain. Thus, βN5-βN6 helps construct the dimer interface.

There are two conserved sequence motifs present in the N-terminal domain. The first one is a Thr-h-Gly-Phe-(Asn/Asp)-Glu-X_4-Arg (h: hydrophobic residue) present within the βN1-αN1 loop; the second one is a Leu-X_2-Gly-h-Arg motif found within the βN4-αN4 loop**.** Glu122 on the βN5-βN6 reverse turn is also strongly conserved. These conserved residues are spatially close to each other, and they form a "pocket" that spans the dimer interface on the N-terminal. The anatomy of this pocket has the distinctive features of a ligand-binding pocket.

Since the pocket is spanning the dimer interface of the protein, a two-fold pseudosymmetric ligand may bind to it. Lintner *et al.* hypothesizes that N-terminal ligand-binding pocket of Csa3 may facilitate DNA binding by transmitting a signal that comes with ligand binding.

A comprehensive analysis by Makarova *et al.* shows that DNA-binding wHTH domains, as well as nuclease domains such as REase and HEPN commonly accompany CARF domains (Makarova *et al* 2014). This strengthens the idea that CARF domains are

able to transmit the signal into an "effector" domain. However, the ligands that may bind to CARF domains (including N-terminal domain of Csa3) still remain unknown. Recently, Yan *et al*. published the crystal structure of Csx3 of *Archaeoglobus fulgidus* bound to RNA (Yan *et al.* 2015). In their study, Csx3 is demonstrated to bind Mg^{2+} and RNA in crystal structures and exhibit Mn^{2+}-dependent RNase activity. Yan *et al.* also describe the fold of Csx3 as a "ferredoxin-like" fold (Figure B-2 B). Ferredoxin-like fold (RNA-recognition motif [RRM], or RAMP domain) is observed commonly in CRISPR-*cas* proteins and it is a mixed β-sheet consisting of four antiparallel β-strands in the $\beta_2\beta_3\beta_1\beta_4$ order with two α-helices making right handed turns and running antiparallel to each other [199](Figure B-2 C). However, Ferredoxin Fold differs significantly from the topology of Csx3. Csx3 lacks the four antiparallel β-strands and, instead, harbors a 6-stranded mixed β-sheet where β-strands run sequentially to each other. On the other hand, $\beta_2\alpha_1\beta_3\alpha_2\beta_4$ structure of Csx3 shows notable similarities to the aforementioned $\beta_1\alpha_1\beta_2\alpha_2\beta_3$ topology of the Rossmann fold. This suggests that the ssRNA may be expected to bind to β2, β3 and β4 strands of Csx3 and one of the phosphate groups in the RNA could be held in place with the helix dipole moment and the unsaturated amine group in the N-terminal end of α2. PDB structure of Csx3 bound to ssRNA (PDB ID: 3WZI) verifies this prediction. Phosphate group of the nucleotide N2 is found at the N-terminal end of α2; β3-α2 loop recognizes N2 and N3. β2-α1 and β4-β5 loops form the pockets that accommodate N2 and N3, respectively.

Similarities between Csx3 fold and Rossmann fold raise the question whether Csx3 is a member of the CARF domain superfamily or not. Although Makarova *et al*. did

not identify Csx3 as a member in their CARF domain study; the ligand binding pocket of Csx3, as a result of its dimeric arrangement, is reminiscent of the putative ligand binding pocket of Csa3. For these reasons, we propose that Csx3 is a distant member of the CARF domain superfamily. Structure of a CARF domain protein bound to its ligand has not been reported yet; and Csx3 bound to ssRNA is a critical part of the search for a ligand for CARF domain proteins.

An important detail to remind ourselves at this point is that interference complexes of type III systems cleave ssRNA targets in 6nt intervals. We combine these two pieces of information, and hypothesize that the “mystery ligand”, which activates Csa3’s regulatory function is a 4-6 nt ssRNA fragment.

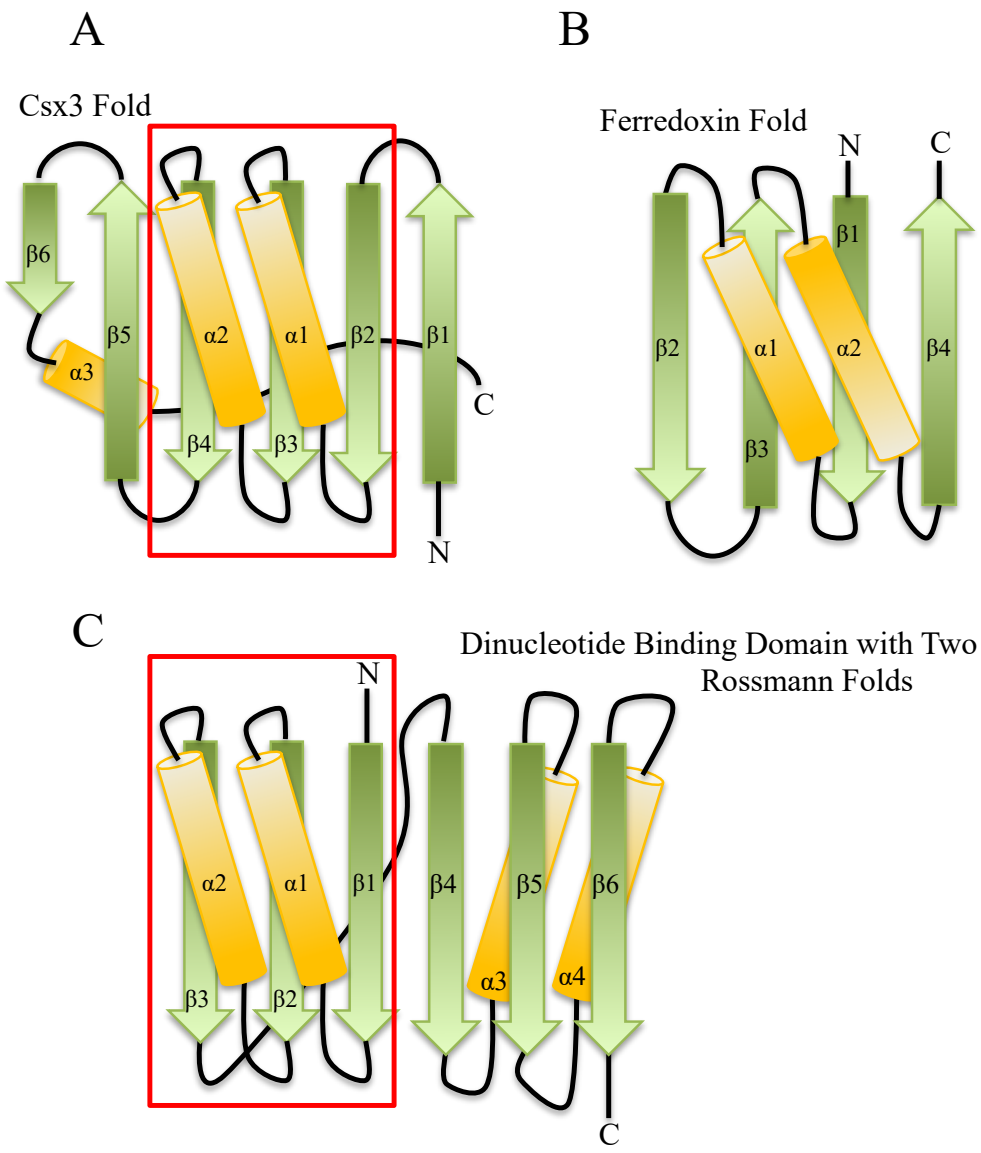


Figure B-2: Topology or fold diagrams of several domains. β strands and α helices are shown in green and yellow, respectively. Color gradients in secondary structures are added to show the N-to-C directionality, morphing from dark to lighter color. Secondary structures are labeled in order of their occurrence in the amino acid sequence. **A)** N-Terminal CARF domain of Csa3 from *S. islandicus.* **B)** Dinucleotide binding domain with two Rossmann Folds. CARF domain loses the α4 connecting β5 and β6, but the α4 of the symmetry related subunit restores this structural element, making β5-β6 reverse turn very important in forming dimer interface. **C)** Ferredoxin fold. It differs significantly from the overall fold of Csx3, as it lacks the 4-antiparallel β-strands of the ferredoxin fold. **D)** Actual fold of Csx3. Figure adapted from Topuzlu et al. (2016).

Sulfolobus islandicus as a model organism for studying regulation of CRISPR-Cas

S. islandicus M16.4, a strain isolated from Mutnovsky Volcano in Kamchatka, Russia, harbors an interrupted *csa3* gene, in which an *att* site whose origins appear to be viral is inserted [200].Another strain, *S. islandicus* M16.27, has 97% identity with M16.4's *csa3* with no viral insertions, and an acquired spacer in *S. islandicus* M16.27 matches perfectly with the viral segment that is inserted in the *S. islandicus* M16.4 *csa3* [193,201].This "viral disruption" of *csa3* in M16.4 prevents it from acquiring new spacers for integration into the two CRISPR loci it harbors (Figure B-3). However, when Csa3 from M16.27 is expressed using an arabinose-inducible promoter in M16.4, M16.4 gains the ability to acquire new spacers in both its CRISPR loci(Figure B-4). In addition to this evolutionary variation between the two strains of *S. islandicus,* a knockout system based on simvastatin selection was developed [201,202]. This system produced *S. islandicus* RJW002 strain, which lacks the ability to produce uracil, due to the complete deletion of its *pyrEF* genes, encoding OPRTase and OMPdecase. Types of *S. islandicus* strains that are mentioned in this paper can be found in Table A-1.

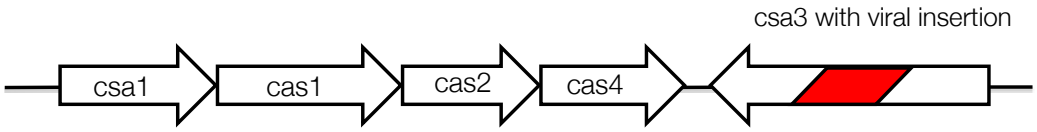


Figure B-3: Acquisition cassette of *S. islandicus* M16.4. *csa3* rests on the opposite strand with respect to *csa1, cas1, cas2* and *cas4.* Viral spacer inserted into *csa3* is shown with the red box.

Table B-1**:** *Sulfolobus islandicus* strains that are mentioned and used in this project.

Strain	**Genotype**	**Phenotype**	**Reference**
M16.4	Wild-type isolated from Mutnovsky Volcano from Kamchatka, Russia. Has a viral insertion in its *csa3* gene.	Cannot acquire new spacers into its CRISPR arrays.	Whitaker et al. 2003
M16.27	Wild-type isolated from the same solfataric field; its CRISPR array contains a spacer which matches 100% with	Able to insert new spacers into its CRISPR arrays.	Reno et al. 2009
RJW002	M 16.4 with *pyrEF* genes completely deleted.	Cannot acquire new spacers, uracil auxotroph.	Zhang et al. 2012
RJW004	M16.4 with *pyrEF* and	Cannot acquire new spacers, uracil and arginine auxotroph.	Zhang et al. 2012

Final remarks

Crystal structure of *si*Csa3 showed a similar ligand binding pocket to *ss*Csa3. We've further attempted to cocrystallize *si*Csa3 with an A4, 4 nucleotide single-strand RNA, as well as a 4 base cyclic RNA. The work related to the cocrystallization efforts are continued by Prof. Martin Lawrence.

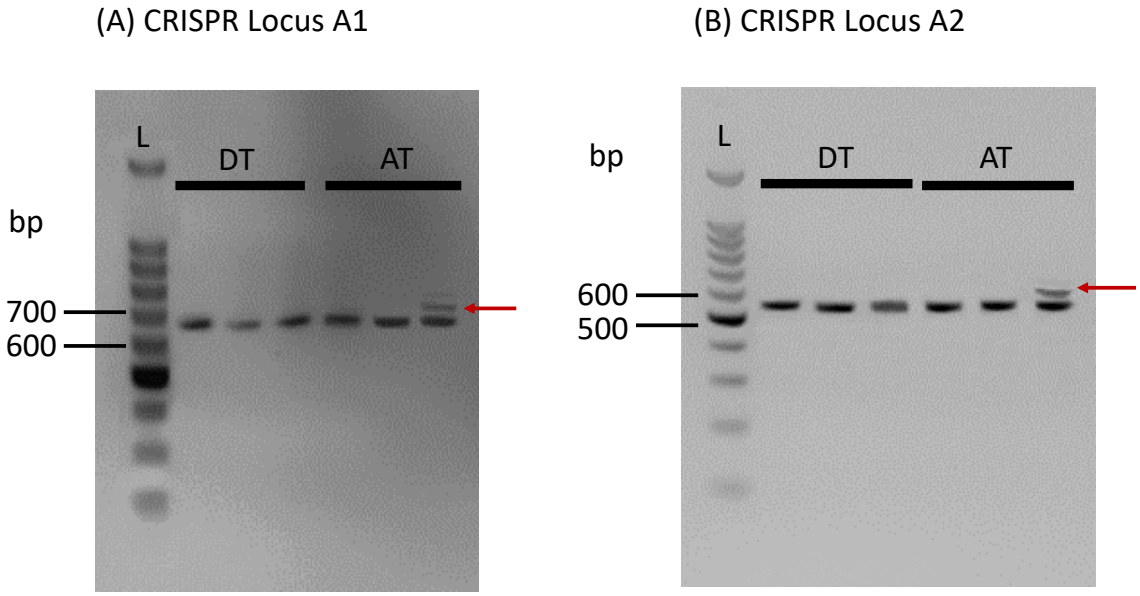


Figure B-4: **(A)** Agarose gel showing spacer acquisition in S. islandicus M16.4 CRISPR Locus A1 in Dextrin (DT; 2-4) or Arabinose (AT; 5-7). Lanes 2&5: S. islandicus M16.4 (RJW004) Lanes 3&6: empty overexpression vector transformed RJW004 [pOE]. Lanes Figure B-4 (continued): 4&7: overexpression vector containing Csa3 transformed RJW004 [pOECsa3]. L: Ladder. Red arrows indicate acquired spacers.
(B) Agarose gel showing spacer acquisition in S. islandicus M16.4 CRISPR Locus A2 in Dextrin (DT; 2-5) or Arabinose (AT; 5-7). Lanes 2&5: S. islandicus M16.4 (RJW004) Lanes 3&6: empty overexpression vector transformed RJW004 [pOE]. Lanes 4&7: overexpression vector containing Csa3 transformed RJW004 [pOECsa3]. L: Ladder. Red arrows indicate acquired spacers.

Materials and Methods

Purification and Expression of Csa3

pSeSD plasmid containing Csa3 gene from *Sulfolobus islandicus* M16.27 was acquired from Rachel Whitaker in University of Illinois. *csa3* from pSeSD was amplified using a 2-step polymerase chain reaction (PCR) in order to prevent primer dimerization. Primers for the first step of the PCR were

ATGCATCACCATCACCATCACATGAAATCATATTTTGTGACC and CACTTTGTACAAGAAAGCTGGGTTCTAGTAAGGTACGTTAATTTCTC, which introduced Shine-Dalgarno and Kozak sequences, and a C-terminal 6xHis tag. Primers for the second step of the PCR were GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAG AACCATGCATCACCATCATCAC and GGGGACCACTTTGTACAAGAAAGCTGGGTCCTA, respectively. ~800bp PCR product was inserted into a pDONR201 vector with a site-specific recombination reaction (Invitrogen) Resulting entry clone was sent for sequencing to confirm there were no mutations (Nevada Genomics). Cloned fragment was transferred to pDEST14 vector with another site-specific recombination reaction and yielded the pEXP14-NHis-M1627_RS4795 expression vector (M1627 is the Sulfolobus species, RS4795 is the gene name in M1627). BL21 (DE3)-pLysS *E. coli* cells were transformed with the expression vector and a single colony was used to grow an overnight culture with 34 μg/ml chloramphenicol and 100 μg/ml ampicillin. ZYP-5052 autoinduction media containing 34 μg/ml chloramphenicol and 100 μg/ml ampicillin was inoculated with thousand-fold diluted overnight culture and grown at 37°C for 14-18 hours. Cells were harvested by centrifugation at 4000***g*** for 15 minutes and pellets were stored in –20°C.

Thawed cell pellets were resuspended at 5 ml of lysis buffer (400 mM NaCl, 20 mM Tris.HCl pH8, 10 mM imidazole and 0.1 mM PMSF) to each gram of cell pellet. Cells were lysed using a microfluidizier and were subsequently incubated at 65°C for 30 minutes for the denaturation of *E. coli* proteins. Lysate was then centrifuged at 20,000***g***

for 45 minutes and supernatant was applied to a gravity-flow column packed with Ni-NTA agarose. Column was washed with the wash buffer (400 mM NaCl, 20 mM Tris.HCl pH8 and 20mM Imidazole) three times with with 8 bed volumes. Protein was eluted off the Ni-NTA column 5 times using lysis buffer (400 mM imidazole, 100mM NaCl, 20 mM Tris.HCl pH8) twice the volume of the Ni-NTA bed. Protein concentration was measured using Bradford Assay. Eluted protein fractions were pooled and dialyzed overnight at 4 oC against a solution of 100 mM NaCl and 10 mM Tris.HCl pH8 in a 3,500 MWCO dialysis membrane (Otherwise protein precipitates within a few hours due to imidazole – get it into dialysis asap). Purity and molecular weight of Csa3 was determined via SDS-PAGE. Dialyzed protein was then concentrated up to 4 mg/ml using a spin concentrator with a 10 kDa cutoff and then applied to a Superdex 75 size-exclusion column equilibrated with 100 mM NaCl and 10 mM Tris.HCl pH8. The predicted size of a single subunit is 27 kDa, but we see a peak at ~ 60 kDa by SEC, suggesting a dimer.

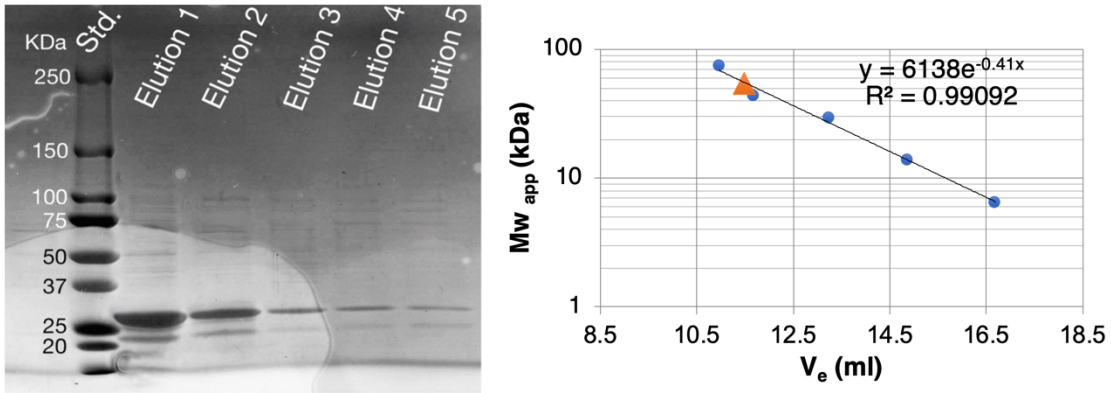


Figure B-5**:** Purification of Csa3 from PlysS cells. **A)** SDS-PAGE gel of the eluted fractions from a Ni-NTA column. **B)** Standard mobility curve constructed for Superdex 75 10/300gl column, using proteins with known molecular weights. Purified Csa3 migrates as a ~60 kDa molecule, suggesting a dimer.

Crystallization

Purified Csa3 was concentrated to 6 mg/ml with a 10,000 MWCO Amicon Ultra™ spin concentrators (Milipore). Protein was crystallized using hanging drop vapor diffusion. Drops were set up at 18°C using 2 µl of concentrated Csa3 and 2 µl of well solution consisting of 7.5% 2-propanol, 20-24% PEG 400 and 0.1 M Piperazine pH 5.0. Csa3 crystals were twinned, multiple crystals emerging from a single nucleation point. These crystals were used as seed stock for microseeding. Twinned crystals obtained from above conditions were used as seed stock in the microseeding experiments. Twinned crystals growing from overnucleated wells were selected and 20 µl mother liquor was placed on a cover slip along with twinned crystals. Crystals were crushed in the drop ad the drop was transferred in a microcentrifuge tube. Tube was then vortexed for 3-5 minutes and used as the “seed stock”. Up to 1:10,000 serial dilutions of the seed stock were made. Drops containing 1µl seed dilution, 2 µl 6 mg/ml Csa3 and 2 µl well solution were set up via sitting drop vapor diffusion. Crystals up to 400µm x 50 µm x 10µm in size were grown in 7-9 days.

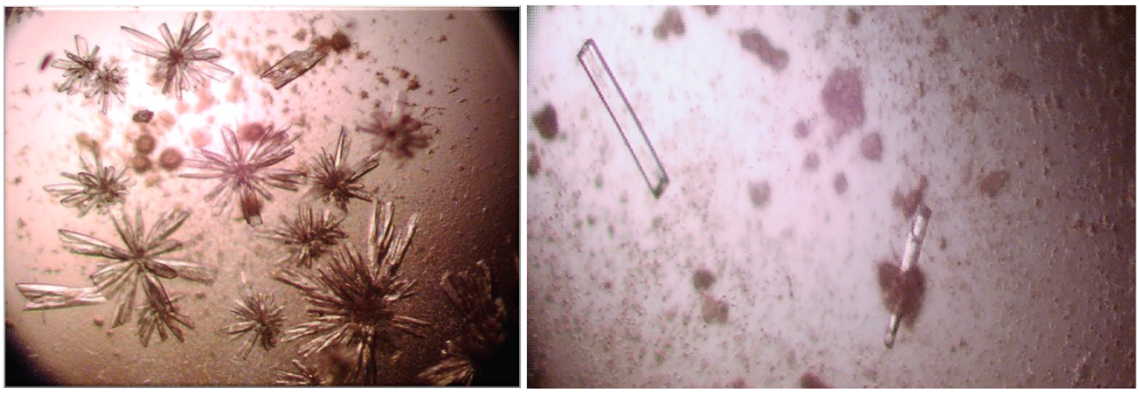


Figure B-6: Left: Multiple Csa3 crystals emerging from a single nucleation point. Right: Single-crystal Csa3.

Data collection and refinement

A diffraction dataset on single-crystal Csa3 was collected Dataset were collected at the home-source R-AXIS IV (Rigaku). Data were indexed and scaled in space group $P2_12_12_1$. Phases were assigned to *si*Csa3 was solved using *ss*Csa3 Chain A (PDB ID: 2WTE, 82% identity) as the model for the PHASER MR pipeline, which placed two copies of *si*Csa3 in the asymmetric unit with 44% solvent content. REFINE was used to obtain an initial refined structure and iterative rounds of refinement were done with COOT.

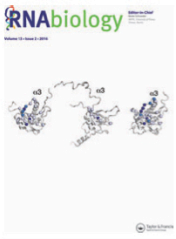
Spacer Acquisition Assays

RJW004/pOE-csa3 and RJW004/pOE strains were grown in 20 mL DT liquid medium without shaking in 25 mL vented cell culture flasks until OD_{600} = 0.2-0.3 cells were harvested and genomic DNA of the strains were extracted, and PCR analysis were run on the genomic DNA template using the following primers tabulated in Table A-2.

Table A-2: Primers used in the spacer acquisition assays.

Primers for A1 locus
16.4A1-sp5-For: 5' GATTATATCAGCAAAAATTACACG 3'
16.4A1-L-Rev: 5' TAAGGAGTGGGGTAAAGACATT 3’
Amplicon: 673bp;

Primers for A2 locus
16.4A2-L-For: 5' TCTATAACTTAAGAGCTTGTAGCTGA 3'
16.4A2-sp5-Rev: 5' AATGAATGTATTTCAAATTCCAAC 3'
Amplicon: 584bp;



RNA Biology

ISSN: 1547-6286 (Print) 1555-8584 (Online) Journal homepage: http://www.tandfonline.com/loi/krnb20

Recognition of a pseudo-symmetric RNA tetranucleotide by Csx3, a new member of the CRISPR associated Rossmann fold superfamily

Ece Topuzlu & C. Martin Lawrence

To cite this article: Ece Topuzlu & C. Martin Lawrence (2016) Recognition of a pseudo-symmetric RNA tetranucleotide by Csx3, a new member of the CRISPR associated Rossmann fold superfamily , RNA Biology, 13:2, 254-257, DOI: 10.1080/15476286.2015.1130209

To link to this article: http://dx.doi.org/10.1080/15476286.2015.1130209

Accepted author version posted online: 04 Jan 2016.
Published online: 04 Jan 2016.

Submit your article to this journal

Article views: 171

View related articles

View Crossmark data

Figure B-7: Recognition of a pseudo-symmetric RNA tetranucleotide by Csx3, a new member of the CRISPR-associated Rossmann fold Published work of Csx3 can be found in RNA biology.

REFERENCES CITED

1. Yoshida, S. *et al.* A Bacterium that degrades and assimilates poly(ethylene terephthalate). *Science (80-.).* **353**, 759 (2016).

2. Austin, H. P. *et al.* Characterization and engineering of a plastic-degrading aromatic polyesterase. *Proc. Natl. Acad. Sci.* 201718804 (2018). doi:10.1073/pnas.1718804115

3. Halden, R. U. Plastics and Health Risks. *Annu. Rev. Public Health* **31**, 179–194 (2010).

4. Thompson, R. C., Swan, S. H., Moore, C. J. & Vom Saal, F. S. Our plastic age. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 1973–1976 (2009).

5. Browne, M. A. Classify plastic waste as hazardous. *Nature* **494**, 169–171 (2013).

6. PlasticsEurope. Annual Review 2017-2018. *Plast. Annu. Rev. 2017-2018* (2018).

7. Geyer, R., Jambeck, J. R. & Law, K. L. Production, use, and fate of all plastics ever made. *Sci. Adv.* **3**, e1700782 (2017).

8. Gourmelon, G. Global Plastic Production Rises, Recycling Lags. *Vital Signs, Worldwatch Inst.* (2015). doi:2244-775X

9. D'Emilio, F. Pope decries 'emergency'; of plastics blighting world's seas. *AP News* (2018). Available at: https://www.apnews.com/9c85604728bd4065a9d759b811b6da92. (Accessed: 28th May 2019)

10. Whinfield, J. R. & Dickson, J. T. Improvements relating to the manufacture of highly polymeric substances. *Br. Pat.* (1941).

11. Ji, L. N. Study on Preparation Process and Properties of Polyethylene Terephthalate (PET). *Appl. Mech. Mater.* (2013). doi:10.4028/www.scientific.net/amm.312.406

12. Todsen, W. L. ChemDoodle 6.0. *J. Chem. Inf. Model.* **54**, 2391–2393 (2014).

13. Moore, W. R. & Sheldon, R. P. The crystallization of polyethylene terephthalate by organic liquids. *Polymer (Guildf).* **2**, 315–321 (1961).

14. Kuzmanović, M. *et al.* Development of crystalline morphology and its relationship with mechanical properties of PP/PET microfibrillar composites containing POE and POE-g-MA. *Polymers (Basel).* **10**, (2018).

15. Wyeth, N. C. Biaxially oriented poly(ethylene terephthalate)bottle. *United Pat. Off.* (1970).

16. Jankauskaite, V., Macijauskas, G. & Lygaitis, R. Polyethylene Terephthalate Waste Recycling and Application Possibilities : a Review. *Mater. Sci.* **14**, (2009).

17. Leian Bartolome, Muhammad Imran, Bong Gyoo Cho, W. A. A.-M. and D. H. K. *Recent Developments in the Chemical Recycling of PET*. *Trends and Perspectives, Dimitris S. Achilias, IntechOpen* (2012). doi:10.5772/33800

18. Achilias, D. S. & Karayannidis, G. P. The chemical recycling of PET in the framework of sustainable development. *Water, Air, Soil Pollut. Focus* (2004). doi:10.1023/B:WAFO.0000044812.47185.0f

19. López-Fonseca, R., González-Marcos, M. P., González-Velasco, J. R. & Gutiérrez-Ortiz, J. I. A kinetic study of the depolymerisation of poly(ethylene terephthalate): By phase transfer catalysed alkaline hydrolysis. *J. Chem. Technol. Biotechnol.* (2009). doi:10.1002/jctb.2011

20. Lithner, D., Larsson, A. & Dave, G. Environmental and health hazard ranking and assessment of plastic polymers based on chemical composition. *Sci. Total Environ.* **409**, 3309–3324 (2011).

21. Jambeck, J. R. *et al.* Plastic waste inputs from land into the ocean. *Science (80-.).* **347**, 768–771 (2015).

22. Acevedo, E. Caretta caretta turtle. 198 (2019).

23. Bergmann, M., Sandhop, N., Schewe, I. & D'Hert, D. Observations of floating anthropogenic litter in the Barents Sea and Fram Strait, Arctic. *Polar Biol.* **39**, 553–560 (2016).

24. Zalasiewicz, J. *et al.* The geological cycle of plastics and their use as a stratigraphic indicator of the Anthropocene. *Anthropocene* **13**, 4–17 (2016).

25. Worm, B., Lotze, H. K., Jubinville, I., Wilcox, C. & Jambeck, J. Plastic as a Persistent Marine Pollutant. *Annu. Rev. Environ. Resour.* **42**, 1–26 (2017).

26. Cox, K. D. *et al.* Human Consumption of Microplastics. *Environ. Sci. Technol.* **53**, acs.est.9b01517 (2019).

27. Peixoto, D. *et al.* Microplastic pollution in commercial salt for human consumption: A review. *Estuar. Coast. Shelf Sci.* **219**, 161–168 (2019).

28. Wright, S. L. & Kelly, F. J. Plastic and Human Health: A Micro Issue? *Environ. Sci. Technol.* **51**, 6634–6647 (2017).

29. Müller, R. J., Kleeberg, I. & Deckwer, W. D. Biodegradation of polyesters containing aromatic constituents. *J. Biotechnol.* **86**, 87–95 (2001).

30. Napper, I. E. & Thompson, R. C. Environmental Deterioration of Biodegradable, Oxo-biodegradable, Compostable, and Conventional Plastic Carrier Bags in the Sea, Soil, and Open-Air over a 3-Year Period. *Environ. Sci. Technol.* **53**, 4775–4783 (2019).

31. Royer, S. J., Ferrón, S., Wilson, S. T. & Karl, D. M. Production of methane and ethylene from plastic in the environment. *PLoS One* **13**, 1553–1555 (2018).

32. Hiç, C., Pradhan, P., Rybski, D. & Kropp, J. P. Food Surplus and Its Climate Burdens. *Environ. Sci. Technol.* **50**, 4269–4277 (2016).

33. Xanthos, D. & Walker, T. R. International policies to reduce plastic marine pollution from single-use plastics (plastic bags and microbeads): A review. *Mar. Pollut. Bull.* **118**, 17–26 (2017).

34. Singh, N. *et al.* Recycling of plastic solid waste: A state of art review and future applications. *Compos. Part B Eng.* **115**, 409–422 (2017).

35. Postconsumer PET Container Recycling Activity in 2015. (2016). Available at: https://www.plasticsrecycling.org/images/pdf/resources/reports/APR_NAPCOR_2015RateReportFINAL.pdf.

36. Craighill, A. L. & Powell, J. C. Lifecycle assessment and economic evaluation of recycling: A case study. *Resour. Conserv. Recycl.* (1996). doi:10.1016/0921-3449(96)01105-6

37. Karayannidis, G. P. & Achilias, D. S. Chemical recycling of poly(ethylene terephthalate). *Macromol. Mater. Eng.* **292**, 128–146 (2007).

38. Hong, M. & Chen, E. Y. X. Chemically recyclable polymers: A circular economy approach to sustainability. *Green Chem.* **19**, 3692–3706 (2017).

39. Pieroni, M. P. P., McAloone, T. C. & Pigosso, D. C. A. Business model innovation for circular economy and sustainability: A review of approaches. *J. Clean. Prod.* **215**, 198–216 (2019).

40. Chen, L., Pelton, R. E. O. & Smith, T. M. Comparative life cycle assessment of fossil and bio-based polyethylene terephthalate (PET) bottles. *J. Clean. Prod.* **137**, 667–676 (2016).

41. Tsiropoulos, I. *et al.* Life cycle impact assessment of bio-based plastics from sugarcane ethanol. *J. Clean. Prod.* **90**, 114–127 (2015).

42. Haider, T. P., Völker, C., Kramm, J., Landfester, K. & Wurm, F. R. Plastics of the Future? The Impact of Biodegradable Polymers on the Environment and on Society. *Angew. Chemie - Int. Ed.* **58**, 50–62 (2019).

43. Hu, H. *et al.* Bio-based poly(butylene 2,5-furandicarboxylate)-b-poly(ethylene glycol) copolymers with adjustable degradation rate and mechanical properties: Synthesis and characterization. *Eur. Polym. J.* **106**, 42–52 (2018).

44. Mehta, R., Kumar, V., Bhunia, H. & Upadhyay, S. N. Synthesis of poly(lactic acid): A review. *J. Macromol. Sci. - Polym. Rev.* **45**, 325–349 (2005).

45. Anjum, A. *et al.* Microbial production of polyhydroxyalkanoates (PHAs) and its copolymers: A review of recent advancements. *Int. J. Biol. Macromol.* **89**, 161–174 (2016).

46. Labet, M. & Thielemans, W. Synthesis of polycaprolactone: A review. *Chem. Soc. Rev.* **38**, 3484–3504 (2009).

47. Gigli, M. *et al.* Poly(butylene succinate)-based polyesters for biomedical applications: A review in memory of our beloved colleague and friend Dr. Lara Finelli. *Eur. Polym. J.* **75**, 431–460 (2016).

48. Ferreira, F. V., Cividanes, L. S., Gouveia, R. F. & Lona, L. M. F. An overview on properties and applications of poly(butylene adipate-co-terephthalate)–PBAT based composites. *Polym. Eng. Sci.* **59**, E7–E15 (2019).

49. Tokiwa, Y., Calabia, B. P., Ugwu, C. U. & Aiba, S. Biodegradability of plastics. *Int. J. Mol. Sci.* **10**, 3722–3742 (2009).

50. Ribitsch, D. *et al.* A new esterase from Thermobifida halotolerans hydrolyses polyethylene terephthalate (PET) and polylactic acid (PLA). *Polymers (Basel).* **4**, 617–629 (2012).

51. 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.* **26**, 1400–1405 (2005).

52. Oda, M. *et al.* Enzymatic hydrolysis of PET: functional roles of three Ca 2+ ions bound to a cutinase-like enzyme, Cut190*, and its engineering for improved activity. *Appl. Microbiol. Biotechnol.* **102**, 10067–10077 (2018).

53. Kawai, F. *et al.* Enzymatic Hydrophilization of Polyester Fabrics Using a Recombinant Cutinase Cut 190 and Their Surface Characterization. *J. Fiber Sci. Technol.* **73**, 8–18 (2017).

54. Mindrebo, J. T., Nartey, C. M., Seto, Y., Burkart, M. D. & Noel, J. P. Unveiling the functional diversity of the Alpha-Beta hydrolase fold in plants. *Curr. Opin. Struct. Biol.* **41**, 233–246 (2016).

55. Qian, Z., Fields, C. J., Yu, Y. & Lutz, S. Recent progress in engineering α/β

hydrolase-fold family members. *Biotechnol. J.* **2**, 192–200 (2007).

56. Verger, R. 'Interfacial activation' of lipases: Facts and artifacts. *Trends Biotechnol.* **15**, 32–38 (1997).

57. Khan, F. I. *et al.* The Lid Domain in Lipases: Structural and Functional Determinant of Enzymatic Properties. *Front. Bioeng. Biotechnol.* **5**, 1–13 (2017).

58. Banoth, L., Singh, M., Tekewe, A. & Banerjee, U. C. Increased enantioselectivity of lipase in the transesterification of dl-(±)-3-phenyllactic acid in ionic liquids. *Biocatal. Biotransformation* **27**, 263–270 (2009).

59. Andreeva, A., Howorth, D., Chothia, C., Kulesha, E. & Murzin, A. G. SCOP2 prototype: A new approach to protein structure mining. *Nucleic Acids Res.* **42**, 310–314 (2014).

60. Liu, Z. *et al.* Structural and Functional Studies of A. oryzae Cutinase: Enhanced Thermostability and Hydrolytic Activity of Synthetic Ester and Polyester Degradation. *J Am Chem Soc.* **131**, 15711–15716 (2009).

61. Derewenda, U. *et al.* Current progress, in crystallographic studies of new upases from filamentous fungi. *Protein Eng. Des. Sel.* **7**, 551–557 (1994).

62. Ronkvist, Å. M., Xie, W., Lu, W. & Gross, R. A. Cutinase-Catalyzed hydrolysis of poly(ethylene terephthalate). *Macromolecules* **42**, 5128–5138 (2009).

63. Thumarat, U. *et al.* Comparison of genetic structures and biochemical properties of tandem cutinase-type polyesterases from Thermobifida alba AHK119. *J. Biosci. Bioeng.* **120**, 491–497 (2015).

64. Perz, V. *et al.* Hydrolysis of synthetic polyesters by Clostridium botulinum esterases. *Biotechnol. Bioeng.* **113**, 1024–1034 (2016).

65. Sasoh, M. *et al.* Characterization of the Terephthalate Degradation Genes of Comamonas sp. Strain E6. *Appl. Environ. Microbiol.* **72**, 1825–1832 (2006).

66. Joo, S. *et al.* Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation. *Nat. Commun.* **9**, (2018).

67. Han, X. *et al.* Structural insight into catalytic mechanism of PET hydrolase. *Nat. Commun.* **8**, 2106 (2017).

68. Fecker, T. *et al.* Active Site Flexibility as a Hallmark for Efficient PET Degradation by I. sakaiensis PETase. *Biophys. J.* **114**, 1302–1312 (2018).

69. Liu, B. *et al.* Protein crystallography and site-direct mutagenesis analysis of the

poly(Ethylene terephthalate) hydrolase petase from Ideonella sakaiensis. *ChemBioChem* **19**, 1471–1475 (2018).

70. Chen, C. C., Han, X., Ko, T. P., Liu, W. & Guo, R. T. Structural studies reveal the molecular mechanism of PETase. *FEBS J.* **285**, 3717–3723 (2018).

71. Chen, C. C., Han, X., Ko, T. P., Liu, W. & Guo, R. T. Structural studies reveal the molecular mechanism of PETase. *FEBS J.* **285**, 3717–3723 (2018).

72. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).

73. Kawai, F., Kawabata, T. & Oda, M. Current knowledge on enzymatic PET degradation and its possible application to waste stream management and other fields. *Appl. Microbiol. Biotechnol.* **103**, 4253–4268 (2019).

74. Shirke, A. N. *et al.* Stabilizing Leaf and Branch Compost Cutinase (LCC) with Glycosylation: Mechanism and Effect on PET Hydrolysis. *Biochemistry* **57**, 1190–1200 (2018).

75. Brunecky, R. *et al.* Revealing Nature's Cellulase Diversity: The Digestion Mechanism of Caldicellulosiroptor bescii CelA. *Science* **342**, 1513–1516 (2013).

76. Palm, G. J. *et al.* Structure of the plastic-degrading Ideonella sakaiensis MHETase bound to a substrate. *Nat. Commun.* **10**, 1–10 (2019).

77. Fuchs, G., Boll, M. & Heider, J. Microbial degradation of aromatic compounds — from one strategy to four. *Nat. Rev. Microbiol.* **9**, 803–816 (2011).

78. Harwood, C. S. & Parales, R. E. the B-Ketoadipate Pathway and the Biology of Self-Identity. *Annu. Rev. Microbiol.* **50**, 553–590 (1996).

79. Xie, N. Z., Liang, H., Huang, R. B. & Xu, P. Biotechnological production of muconic acid: Current status and future prospects. *Biotechnol. Adv.* **32**, 615–622 (2014).

80. Rorrer, N. A. *et al.* Renewable Unsaturated Polyesters from Muconic Acid. *ACS Sustain. Chem. Eng.* **4**, 6867–6876 (2016).

81. Suastegui, M. *et al.* Combining Metabolic Engineering and Electrocatalysis: Application to the Production of Polyamides from Sugar. *Angew. Chemie - Int. Ed.* **55**, 2368–2373 (2016).

82. Vardon, D. R. *et al.* Adipic acid production from lignin. *Energy Environ. Sci.* **8**, 617–628 (2015).

83. Zeikus, J. G., Jain, M. K. & Elankovan, P. Biotechnology of succinic acid

production and markets for derived industrial products. *Appl. Microbiol. Biotechnol.* **51**, 545–552 (1999).

84. Lu, R. *et al.* Production of Diethyl Terephthalate from Biomass-Derived Muconic Acid. *Angew. Chemie - Int. Ed.* **55**, 249–253 (2016).

85. Rorrer, N. A., Vardon, D. R., Dorgan, J. R., Gjersing, E. J. & Beckham, G. T. Biomass-derived monomers for performance-differentiated fiber reinforced polymer composites. *Green Chem.* **19**, 2812–2825 (2017).

86. Johnson, C. W. *et al.* Innovative Chemicals and Materials from Bacterial Aromatic Catabolic Pathways. *Joule* **3**, 1523–1537 (2019).

87. Berrow, N. S. *et al.* Recombinant protein expression and solubility screening in Escherichia coli: A comparative study. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 1218–1226 (2006).

88. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. **41**, 207–234 (2005).

89. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).

90. Payne, C. M. *et al.* Fungal cellulases. *Chem. Rev.* **115**, 1308–1448 (2015).

91. Chanzy, H. & Henrissat, B. Undirectional degradation of Valonia cellulose microcrystals subjected to cellulase action Trichoderma reesei Valonia Endo-exo cooperation Cellulose chain polarity. **184**, 0–3 (2573).

92. Boisset, C., Fraschini, C., Schülein, M., Henrissat, B. & Chanzy, H. Imaging the enzymatic digestion of bacterial cellulose ribbons reveals the endo character of the cellobiohydrolase Cel6A from Humicola insolens and its mode of synergy with cellobiohydrolase Cel7A. *Appl. Environ. Microbiol.* **66**, 1444–1452 (2000).

93. Haubruge, H. G., Jonas, A. M. & Legras, R. Staining of poly(ethylene terephthalate) by ruthenium tetroxide. *Polymer (Guildf).* **44**, 3229–3234 (2003).

94. Jorda, J. & Yeates, T. O. Widespread disulfide bonding in proteins from thermophilic archaea. *Archaea* **2011**, (2011).

95. Chan, M. K., Mukund, S., Kletzin, A., Adams, M. W. & Rees, D. C. Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science (80-.).* **267**, 1463–1469 (1995).

96. Yip, K. S. P. *et al.* Insights into the molecular basis of thermal stability from the

analysis of ion-pair networks in the glutamate dehydrogenase family. *Eur. J. Biochem.* **255**, 336–346 (1998).

97. Thompson, M. J. & Eisenberg, D. Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. *J. Mol. Biol.* **290**, 595–604 (1999).

98. Rainer, J. Protein stability and molecular adaptation to extreme conditons. *Eur. J. Biochem.* **202**, 715–728 (1991).

99. Lintner, N. G. *et al.* The structure of the CRISPR-associated protein csa3 provides insight into the regulation of the CRISPR/Cas system. *J. Mol. Biol.* **405**, 939–955 (2011).

100. Topuzlu, E. & Lawrence, C. M. Recognition of a Pseudo-Symmetric RNA Tetranucleotide by Csx3, a New Member of the CRISPR Associated Rossmann Fold Superfamily. *RNA Biol.* **6286**, 00–00 (2016).

101. Then, J. *et al.* A disulfide bridge in the calcium binding site of a polyester hydrolase increases its thermal stability and activity against polyethylene terephthalate. *FEBS Open Bio* **6**, 425–432 (2016).

102. Gonçalves, A. M., Serro, A. P., Aires-Barros, M. R. & Cabral, J. M. S. Effects of ionic surfactants used in reversed micelles on cutinase activity and stability. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1480**, 92–106 (2000).

103. Teresa Cunha, M., Cabral, J. M. S., Tjerneld, F. & Raquel Aires-Barros, M. Effect of salts and surfactants on the partitioning of fusarium solani pisi cutinase in aqueous two-phase systems of thermoseparating ethylene oxide/propylene oxide random copolymer and hydroxypropyl starch. *Bioseparation* **9**, 203–209 (2000).

104. Lindsay, R. J., Johnson, Q. R., Evangelista, W., Nellas, R. B. & Shen, T. DMSO enhanced conformational switch of an interfacial enzyme. *Biopolymers* 864–872 (2016). doi:10.1002/bip.22924

105. Bubner, P., Plank, H. & Nidetzky, B. Visualizing cellulase activity. *Biotechnol. Bioeng.* **110**, 1529–1549 (2013).

106. Law, K. L. *et al.* Plastic accumulation in the North Atlantic subtropical gyre. *Science (80-.).* **329**, 1185–1188 (2010).

107. Cózar, A. *et al.* Plastic debris in the open ocean. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 10239–10244 (2014).

108. Jambeck, J. R. *et al.* Plastic waste inputs from land into the ocean. *Science (80-.).* **347**, 768–771 (2015).

109. Fiorenza, E. A. *et al.* Plastic waste associated with disease on coral reefs. *Science (80-.).* **359**, 460–462 (2018).

110. de Souza Machado, A. A., Kloas, W., Zarfl, C., Hempel, S. & Rillig, M. C. Microplastics as an emerging threat to terrestrial ecosystems. *Glob. Chang. Biol.* **24**, 1405–1416 (2018).

111. Allen, S. *et al.* Atmospheric transport and deposition of microplastics in a remote mountain catchment. *Nat. Geosci.* **12**, 339–344 (2019).

112. Restrepo-Flórez, J. M., Bassi, A. & Thompson, M. R. Microbial degradation and deterioration of polyethylene - A review. *Int. Biodeterior. Biodegrad.* **88**, 83–90 (2014).

113. Yang, J., Yang, Y., Wu, W. M., Zhao, J. & Jiang, L. Evidence of polyethylene biodegradation by bacterial strains from the guts of plastic-eating waxworms. *Environ. Sci. Technol.* **48**, 13776–13784 (2014).

114. Yang, Y. *et al.* Biodegradation and Mineralization of Polystyrene by Plastic-Eating Mealworms: Part 2. Role of Gut Microorganisms. *Environ. Sci. Technol.* **49**, 12087–12093 (2015).

115. Bombelli, P., Howe, C. J. & Bertocchini, F. Polyethylene bio-degradation by caterpillars of the wax moth Galleria mellonella. *Curr. Biol.* **27**, R292–R293 (2017).

116. Dvořák, P., Nikel, P. I., Damborský, J. & de Lorenzo, V. Bioremediation 3.0: Engineering pollutant-removing bacteria in the times of systemic biology. *Biotechnol. Adv.* **35**, 845–866 (2017).

117. Wierckx, N. *et al.* Plastic waste as a novel substrate for industrial biotechnology. *Microb. Biotechnol.* **8**, 900–903 (2015).

118. Wei, R. & Zimmermann, W. Biocatalysis as a green route for recycling the recalcitrant plastic polyethylene terephthalate. *Microb. Biotechnol.* **10**, 1302–1307 (2017).

119. Wei, R. & Zimmermann, W. Microbial enzymes for the recycling of recalcitrant petroleum-based plastics: how far are we? *Microb. Biotechnol.* **10**, 1308–1322 (2017).

120. Taniguchi, I. *et al.* Biodegradation of PET: Current Status and Application Aspects. *ACS Catal.* 4089–4105 (2019). doi:10.1021/acscatal.8b05171

121. Araújo, R. *et al.* Tailoring cutinase activity towards polyethylene terephthalate and polyamide 6,6 fibers. *J. Biotechnol.* **128**, 849–857 (2007).

122. de Castro, A. M., Carniel, A., Nicomedes Junior, J., da Conceição Gomes, A. & Valoni, É. Screening of commercial enzymes for poly(ethylene terephthalate) (PET) hydrolysis and synergy studies on different substrate sources. *J. Ind. Microbiol. Biotechnol.* **44**, 835–844 (2017).

123. Herrero Acero, E. *et al.* Enzymatic surface hydrolysis of PET: Effect of structural diversity on kinetic properties of cutinases from Thermobifida. *Macromolecules* **44**, 4632–4640 (2011).

124. Ribitsch, D. *et al.* Characterization of a new cutinase from Thermobifida alba for PET-surface hydrolysis. *Biocatal. Biotransformation* **30**, 2–9 (2012).

125. Przylucka, A. *et al.* Enhanced Cutinase-Catalyzed Hydrolysis of Polyethylene Terephthalate by Covalent Fusion to Hydrophobins. *Appl. Environ. Microbiol.* **81**, 3586–3592 (2015).

126. Roth, C. *et al.* Structural and functional studies on a thermostable polyethylene terephthalate degrading hydrolase from Thermobifida fusca. *Appl. Microbiol. Biotechnol.* **98**, 7815–7823 (2014).

127. Sulaiman, S. *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.* **78**, 1556–1562 (2012).

128. Furukawa, M., Kawakami, N., Oda, K. & Miyamoto, K. Acceleration of Enzymatic Degradation of Poly(ethylene terephthalate) by Surface Coating with Anionic Surfactants. *ChemSusChem* **11**, 4018–4025 (2018).

129. Ma, Y. *et al.* Enhanced Poly(ethylene terephthalate) Hydrolase Activity by Protein Engineering. *Engineering* **4**, 888–893 (2018).

130. Son, H. F. *et al.* Rational Protein Engineering of Thermo-Stable PETase from Ideonella sakaiensis for Highly Efficient PET Degradation. *ACS Catal.* **9**, 3519–3526 (2019).

131. Rauwerdink, A. & Kazlauskas, R. J. How the Same Core Catalytic Machinery Catalyzes 17 Different Reactions: The Serine-Histidine-Aspartate Catalytic Triad of α/β-Hydrolase Fold Enzymes. *ACS Catal.* **5**, 6153–6176 (2015).

132. Hermoso, J. A. *et al.* The crystal structure of feruloyl esterase A from Aspergillus niger suggests evolutive functional convergence in feruloyl esterase family. *J. Mol. Biol.* **338**, 495–506 (2004).

133. Suzuki, K. *et al.* Crystal structure of a feruloyl esterase belonging to the tannase family: A disulfide bond near a catalytic triad. *Proteins Struct. Funct. Bioinforma.* **82**, 2857–2867 (2014).

134. Mitchell, A. L. *et al.* InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **47**, D351–D360 (2019).

135. Dilokpimol, A. *et al.* Diversity of fungal feruloyl esterases: updated phylogenetic classification, properties, and industrial applications. *Biotechnol. Biofuels* **9**, 1–18 (2016).

136. Barth, M. *et al.* Effect of hydrolysis products on the enzymatic degradation of polyethylene terephthalate nanoparticles by a polyester hydrolase from Thermobifida fusca. *Biochem. Eng. J.* **93**, 222–228 (2015).

137. Neidhardt, F. C., Bloch, P. L. & Smith, D. F. Culture medium for enterobacteria. *Microbiology* **119**, 736–747 (1974).

138. Kabsch, W. {\it XDS}. *Acta Crystallogr. Sect. D* **66**, 125–132 (2010).

139. Collaborative CP (1994) The CCP4 suite: programs for protein crystallography. Acta Cryst. D 50(Pt 5):760.

140. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

141. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 479–485 (2010).

142. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

143. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 352–367 (2012).

144. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).

145. Engh, R. A. & Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. Sect. A* **47**, 392–400 (1991).

146. Capra, J. A. & Singh, M. Predicting functionally important residues from sequence conservation. *Bioinformatics* **23**, 1875–1882 (2007).

147. Cock, P. J. A. *et al.* Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).

148. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).

149. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

150. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

151. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **8**, 275–282 (1992).

152. Nei, M. & Kumar, S. *Molecular Evolution and Phylogenetics*. (Oxford University Press, New York, 1992).

153. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).

154. Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided. Mol. Des.* **27**, 221–234 (2013).

155. Anonymous (Schrödinger Release 2017-3: Schrödinger Suite 2017-3 Protein Preparation Wizard.).

156. Anonymous (Schrödinger Release 2017-3: Impact, Schrödinger, LLC, New York, NY, 2016.).

157. Banks, J. L. *et al.* Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **26**, 1752–1780 (2005).

158. Greenwood, J. R., Calkins, D., Sullivan, A. P. & Shelley, J. C. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J. Comput. Aided. Mol. Des.* **24**, 591–604 (2010).

159. Anonymous (Schrödinger Suite 2018-4 Induced Fit Docking protocol; Glide, Schrödinger, LLC, New York, NY, 2018; Prime, Schrödinger, LLC, New York, NY, 2018.).

160. Farid, R., Day, T., Friesner, R. A. & Pearlstein, R. A. New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorganic Med. Chem.* **14**, 3160–3173 (2006).

161. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A. & Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **49**, 534–553 (2006).

162. Anonymous (Schrödinger Release 2019-1: LigPrep, Schrödinger, LLC, New York, NY, 2019.).

163. Nyyssölä, A. Which properties of cutinases are important for applications? *Appl. Microbiol. Biotechnol.* **99**, 4931–4942 (2015).

164. Gamerith, C. *et al.* Enzymatic recovery of polyester building blocks from polymer blends. *Process Biochem.* **59**, 58–64 (2017).

165. Haubruge, H. G., Jonas, A. M. & Legras, R. Morphological study of melt-crystallized poly(ethylene terephthalate). A. Comparison of transmission electron microscopy and small-angle X-ray scattering of bulk samples. *Macromolecules* **37**, 126–134 (2004).

166. Zumstein, M. T. *et al.* Enzymatic Hydrolysis of Polyester Thin Films at the Nanoscale: Effects of Polyester Structure and Enzyme Active-Site Accessibility. *Environ. Sci. Technol.* **51**, 7476–7485 (2017).

167. Baldenegro-Perez, L. A. *et al.* Molecular weight and crystallization temperature effects on poly(ethylene terephthalate) (PET) homopolymers, an isothermal crystallization analysis. *Polymers (Basel).* **6**, 583–600 (2014).

168. Pucci, F., Kwasigroch, J. M. & Rooman, M. SCooP: an accurate and fast predictor of protein stability curves as a function of temperature. *Bioinformatics* **33**, 3415–3422 (2017).

169. Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome Biol.* **18**, 1–15 (2017).

170. Arnett, D. K. & Claas, S. A. Omics of blood pressure and hypertension. *Circ. Res.* **122**, 1409–1419 (2018).

171. Del Chierico, F. *et al.* Gut microbiota profiling of pediatric nonalcoholic fatty liver disease and obese patients unveiled by an integrated meta-omics-based approach. *Hepatology* **65**, 451–464 (2017).

172. Wheelock, Å. M. & Wheelock, C. E. Trials and tribulations of 'omics data analysis: Assessing quality of SIMCA-based multivariate models using examples from pulmonary medicine. *Mol. Biosyst.* **9**, 2589–2596 (2013).

173. Berger, B., Peng, J. & Singh, M. Computational solutions for omics data. *Nat. Rev. Genet.* **14**, 333 (2013).

174. Tanaka, H. Omics-based Medicine and Systems Pathology. *Methods Inf Med* **49**, 173–185 (2010).

175. Wheelock, C. E. *et al.* Application of 'omics technologies to biomarker discovery in inflammatory lung diseases. *Eur. Respir. J.* **42**, 802–825 (2013).

176. Dale, E., Bang-Andersen, B. & Sánchez, C. Emerging mechanisms and treatments for depression beyond SSRIs and SNRIs. *Biochem. Pharmacol.* **95**, 81–97 (2015).

177. Fuchs, A. L. *et al.* Characterization of the antibacterial activity of Bald's eyesalve against drug resistant Staphylococcus aureus and Pseudomonas aeruginosa. *PLoS One* **13**, 1–17 (2018).

178. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653–663 (2005).

179. Bolotin, A., Quinquis, B., Sorokin, A. & Dusko Ehrlich, S. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).

180. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182 (2005).

181. Grissa, I., Vergnaud, G. & Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**, 172 (2007).

182. Rath, D., Amlinger, L., Rath, A. & Lundgren, M. The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie* **117**, 119–128 (2015).

183. Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **1**, 0474–0483 (2005).

184. Sorek, R., Lawrence, C. M. & Wiedenheft, B. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.* **82**, 237–66 (2013).

185. Makarova, K. S. *et al.* Evolution and Classification of the CRISPR–Cas Systems. *Nat. Rev. Microbiol.* **9**, 467–477 (2011).

186. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).

187. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).

188. Sashital, D. G., Wiedenheft, B. & Doudna, J. A. Mechanism of Foreign DNA Selection in a Bacterial Adaptive Immune System. *Mol. Cell* **46**, 606–615 (2012).

189. Samai, P. *et al.* Co-transcriptional DNA and RNA cleavage during type III CRISPR-cas immunity. *Cell* **161**, 1164–1174 (2015).

190. Osawa, T., Inanaga, H., Sato, C. & Numata, T. Crystal structure of the crispr-cas RNA silencing cmr complex bound to a target analog. *Mol. Cell* **58**, 418–430 (2015).

191. Vestergaard, G., Garrett, R. a & Shah, S. a. CRISPR adaptive immune systems of Archaea. *RNA Biol.* **11**, 156–67 (2014).

192. Shah, S. A., Hansen, N. R. & Garrett, R. A. Distribution of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism. *Biochem. Soc. Trans.* **37**, 23–28 (2009).

193. Shah, S., Erdmann, S., Mojica, F. & Garrett, R. Protospacer recognition motifs. *RNA Biol.* **10**, 891–899 (2013).

194. Manica, A. & Schleper, C. CRISPR-mediated defense mechanisms in the hyperthermophilic archaeal genus Sulfolobus. *RNA Biol.* **10**, 671–8 (2013).

195. Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiol. Rev.* **29**, 231–262 (2005).

196. Makarova, K. S., Anantharaman, V., Grishin, N. V., Koonin, E. V. & Aravind, L. CARF and WYL domains: Ligand-binding regulators of prokaryotic defense systems. *Front. Genet.* **5**, 1–9 (2014).

197. Liu, T. *et al.* Transcriptional regulator-mediated activation of adaptation genes triggers CRISPR de novo spacer acquisition. *Nucleic Acids Res.* **43**, 1044–1055 (2015).

198. Rao, S. T. & Rossmann, M. G. Comparison of super-secondary structures in proteins. *J. Mol. Biol.* **76**, 241–256 (1973).

199. Reeks, J., Naismith, J. H. & White, M. F. CRISPR interference: a structural perspective. *Biochem. J* **453**, 155–166 (2013).

200. Whitaker, R. J., Grogan, D. W. & Taylor, J. W. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science (80-.).* **301**, 976–978 (2003).

201. Zhang, C. & Whitaker, R. J. A broadly applicable gene knockout system for the thermoacidophilic archaeon Sulfolobus islandicus based on simvastatin selection. *Microbiol. (United Kingdom)* **158**, 1513–1522 (2012).

202. Zheng, T. *et al.* Development of a simvastatin selection marker for a hyperthermophilic acidophile, Sulfolobus islandicus. *Appl. Environ. Microbiol.* **78**, 568–574 (2012).