

Exploration of reconstructed PET hydrolase ancestors by biophysical and computational methods

by

UPNEET BALA

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Chemistry

Waterloo, Ontario, Canada 2023

© Upneet Bala 2023

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Ancestral sequence reconstruction (ASR), originally developed for the study of molecular evolution, has gained attention as a promising method for protein engineering in the recent decade. The ancestors revived by this technique have often exhibited valuable traits such as improved stability, activity, and substrate promiscuity. Currently, at each stage of the design many non-trivial user decisions are required, specifically for the curation of the multiple sequence alignment and subsequent analysis. Here, we provide a glimpse into the ancestral sequence reconstruction for PET-hydrolases. *Ideonella sakaiensis* PETase has been acclaimed for its ability to hydrolyze polyethylene terephthalate, and is desired for potential applications to green plastic recycling. The natural enzyme however suffers with low stability in even moderate conditions. We hypothesized by including active and thermostable homologs in the phylogenetic tree, we could probabilistically combine their properties to create a more thermostable and active ancestor. These ancestors were vetted computationally and experimentally using molecular dynamics (MD), circular dichroism (CD), Fourier-transform infrared spectroscopy (FTIR), chemical denaturation, and activity assays. Preliminary results show that the expressed ancestors have signs of co-operative folding and secondary structure, but are only marginally stable, aggregation prone, and appear inactive against generic esterase substrates. These qualities are hypothesized to be in part due to the inclusion of non-consensus PETase loops in the multiple sequence alignment, and impaired C-terminal disulfide formation. These results provide a valuable stepping-stone for the next iteration of design of ancestral PET-hydrolases, and enriches the understanding of ASR outcomes to the larger research community.

Acknowledgements

This research was conducted in the Department of Chemistry of the University of Waterloo under the supervision of Professor Elizabeth Meiering and Professor Subha Kalyaanamoorthy. The author is grateful for the sheer volume of learning that occurred under the present and patient guidance of both brilliant professors. Their motivation, love for details, and abilities to get things done have been adopted in the authors mind, and will remain there always. The author has been captivated by the collaboration and teamwork in the lab and will miss the lovely people, the inventive ideas, and the mirthful science.

The author would further like to thank the University of British Columbia, particularly Dr. Michael Murphy and lab, for their gracious reception of the author and support in protein crystallography. An enormous acknowledgement is given to all the administrative staff that have helped make all of these experiences possible.

Importantly, the author would like to acknowledge NSERC APRENTICE for their funding of the project, and their devoted teaching of protein engineering through classes, seminars, and events.

Finally, the author dedicates this thesis to her beloved family—her inspiring mom, Manpreet, her ingenious dad, Gurpreet, the best brother, Jas, the masterly grandparents, the most loving partner, Andrew Jackson, the happy aunts and uncles, the laughter-filled cousins, and finally, the most solid-gold group of friends. You all have made life so beautiful to live.

Table of Contents

Author's Declaration.....	ii
Abstract.....	iii
Acknowledgements.....	iv
List of Figures.....	vii
List of Tables.....	ix
List of Abbreviations.....	x
1 Introduction.....	1
1.1 Plastic Management and Sustainable Recycling.....	1
1.2 Types of Plastics and Recycling Methods.....	3
1.3 Polyethylene Terephthalate (PET) Degrading Enzymes.....	8
1.3.1 Natural PET Hydrolases.....	8
1.3.2 Engineered PET Hydrolases.....	16
1.4 Ancestral Sequence Reconstruction as a Method for Creating Promiscuous and Stable Proteins.....	18
1.4.1 Introduction to Ancestral Sequence Reconstruction.....	18
1.4.2 Maximum likelihood for Ancestral Sequence Reconstruction.....	19
2 Materials and Methods.....	22
2.1 In Silico.....	22
2.1.1 Generation of Ancestor Sequences.....	22
2.1.2 Homology Modelling and Analysis.....	25
2.1.3 AMBER All-Atom Molecular Dynamics.....	26
2.2 Experimental.....	29
2.2.1 Purification of wildtype Tf. Cut II.....	29
2.2.2 Purification of Ancestor Proteins.....	30
2.2.3 Chemical Denaturation Curves.....	32
2.2.4 Circular Dichroism.....	33
2.2.5 ATR-FTIR.....	34
2.2.6 pNPA and pNPP activity.....	35
2.2.7 Mass Spectrometry.....	36
2.2.8 Iodoacetamide Gel.....	37
3 Results.....	38
3.1 A rich sequence space leads to important, early choices.....	38

3.2	Locating desirable ancestral nodes for reconstruction	41
3.3	Ancestor expression and purification	47
3.4	Chemical denaturation suggests low stability of ancestors	52
3.5	Inferred Structure of Ancestors by CD, and FTIR	57
3.6	Enzymatic Activity of Tf. Cut II	62
3.7	Sequence and structural based analysis of ancestors	65
3.7.1	Building of homology models, and overall predicted variation	65
3.7.2	Sequence analysis of ancestors compared to known PET hydrolases	72
3.7.3	Analysis based on stability and solubility predictors.....	78
3.7.4	Disulfide bond formation.....	79
3.8	Synthesis and Summary of Key Sites	82
4	Discussion.....	86
4.1	Future Ideas	91
4.2	Conclusions	94
	References.....	95
	Appendices.....	110
A.	Engineering of Known PET Hydrolases	110
B.	Cut-PETase Multiple Sequence Alignment.....	137
C.	Cut-Therm Multiple Sequence Alignment.....	142
D.	Alignment of Known PET Hydrolases	147
E.	Phylogenetic Trees.....	149
F.	Mass Spectrometry	152
G.	Location of Tryptophan Residues in Tf. Cut II and Ancestors	155
H.	Temperature denaturation of Tf. Cut II.....	156
I.	Iodoacetamide labelling and gel electrophoresis.....	157
J.	Structural comparison of ancestor homology models to Tf. Cut II and Is. PETase	158
K.	Summary for Models.....	159
L.	Key data related to Ancestor Molecular Dynamics	161
M.	ANC Cut-PETase Extended Residue Data.....	164
N.	ANC Cut-Therm Extended Residue Data	174

List of Figures

Figure 1.1 Relative production of plastics by type and recycling options.....	5
Figure 1.2 Phylogenetic relationships between natural plastic degrading enzymes.....	7
Figure 1.3 Suggested catabolic pathway for PET degradation to oxaloacetate and pyruvate.	12
Figure 1.4 Structural characteristics of PET hydrolases.....	15
Figure 1.5 Example of Phylogenetic Tree to demonstrate main ideas in ASR.....	20
Figure 3.1 Sequences of ancestors Cut-PETase and Cut-Therm.....	43
Figure 3.2 Cut-PETase Phylogenetic Tree.....	44
Figure 3.3 Cut-Therm Phylogenetic Tree	46
Figure 3.4 Expression of Tf. Cut II, ANC Cut-Therm, and ANC Cut-PETase.....	48
Figure 3.5 Chemical denaturation reveals native baseline for ancestors, but lower thermodynamic stability compared to Tf. Cut II.....	53
Figure 3.6 CD spectra for Tf. Cut II, ANC Cut-PETase, ANC Cut-Therm. (Top)	59
Figure 3.7 FTIR spectra for Tf. Cut II, ANC Cut-PETase, ANC Cut-Therm.	61
Figure 3.8 Hydrolysis of pNPA with Tf. Cut II.....	64
Figure 3.9 Presence of loop insert #1 is coupled with gain of P118 and loss of conserved Arg trio.	68
Figure 3.10 Loop insert #2 is implicated in active site coordination.....	71
Figure 3.11 Comparison of ancestor sequences to literature mutations previously tested on a variety of PET hydrolase enzymes.	77
Figure 3.12 Molecular dynamics reveal flexible C-terminal may adversely impact disulfide formation.....	81
Figure 3.13 Synthesis of results from simulations, pocket analysis, and comparison to Thh Est may aid in determining critical ANC sites.....	85
Figure 4.1 Sequence based analysis shows relative relationships between PET hydrolases and sequence similarity in buried regions.	93
Figure A1 Multiple Sequence Alignment for Cut-PETase.....	141
Figure A2 Multiple Sequence Alignment for Cut-Therm.....	146
Figure A3 Alignments of known PET hydrolases from Table 1 and ancestors.....	148
Figure A4 pI Representation of Cut-Therm Phylogenetic Tree.....	150
Figure A5 pI Representation of Cut-PETase Phylogenetic Tree.....	151
Figure A6 MALDI mass spectrometry analysis of ANC Cut-PETase.	152
Figure A7 MALDI mass spectrometry of protein standards.	153
Figure A8 Standard curve relating MALDI m/z to molecular weight (Da).	154
Figure A9 Location of Trp Residues on Tf. Cut II and Ancestors	155
Figure A10 Increase in Tf. Cut II light scattering with temperature may correlate to Tm.	156
Figure A11 Iodo-acetamide labelling of post-chemical denaturation samples reveal ancestor disulfides were mainly reduced.	157
Figure A12 Deviation of predicted ancestor homology models from Tf. Cut II and Is. PETase structures.....	158

Figure A13 RMSF for all top-scoring homology models for ANC Cut-PETase from several programs.	161
Figure A14 Changes to active site loop may affect catalytic triad for ANC Cut-Therm more than ANC Cut-PETase.....	163

List of Tables

Table 1. Characterized Natural PET Hydrolysing Enzymes.	9
Table 2 BLASTp hits for Tf. Cut II	22
Table 3 Annotation based separation of remaining BLASTp candidates.....	24
Table 4 Tryptophan fluorescence and light scattering as a metrics to determine optimal refolding buffer with respect to folding and aggregation	50
Table 5 Predicted secondary structure content from BeStSel.....	60
Table 6 Changes to ANC Cut-Therm and Cut-PETase sequence with respect to closet sequence match, Thh_Est.	74
Table 7 Predictions of stability and solubility by ProtParam, LRO/ACO, and AGGRESCAN/Camsol.....	78
Table A1 Literature Survey of PET Hydrolase Variants.....	110
Table A2 MolProbity, Fold-X, and Rosetta Scores for Ancestor Homology Models	159
Table A3 Extended residue data for ANC Cut-PETase.....	164
Table A4 Extended Residue Data for ANC Cut-Therm.	174

List of Abbreviations

2TY: Yeast Extract Tryptone Media

ACO: Absolute Contact Order

ASR: Ancestral Sequence Reconstruction

BeStSel: Beta Structure Selection

BHET: Bis(2-hydroxyethyl) Terephthalate

CD: Circular Dichroism

DMSO: Dimethyl Sulfoxide

EDTA: Ethylenediaminetetraacetic Acid

FTIR: Fourier-transform Infrared Spectroscopy

GAPSEQs: Pertaining to sequences in the alignment missing residues in Loop #1 (residues 111-113) and in Loop #2 (residues 221-223). This is shown well in APPENDIX D.

INSEQs: Pertaining to sequences in the alignment with residues present in Loop #1 (residues 111-113) and Loop #2 (residues 221-223). This is shown well in APPENDIX D.

IPTG: Isopropyl β -D-1-thiogalactopyranoside

LRO: Long-Range Order

MHET: 4-Methyl-5-Hydroxyethylthiazole

PET: Polyethylene Terephthalate

pNPA: p-Nitrophenyl Acetate

pNPP: p-Nitrophenyl Palmitate

RMSF: Root Mean Square Fluctuation

RMSD: Root Mean Square Deviation

SASA: Solvent-Accessible Surface Area

SDS-PAGE: Sodium Dodecyl-Sulfate Polyacrylamide Gel Electrophoresis

TCA: Tricarboxylic Acid Cycle, Krebs Cycle, or Citric Acid Cycle

TPA: Terephthalic Acid

1 Introduction

1.1 Plastic Management and Sustainable Recycling

The mass production of plastics began only in the 1950s following its propitious use in World War II (Leal Filho, Saari, ... Voronova, 2019; PlasticsEurope, 2022). Due to advantageous properties such as flexibility, impermeability, and recalcitrance as well as low cost/ease in manufacturing, plastic polymers have saturated many global industries. Plastic management norms are fragmented, and the intensity of protocols in place varies nationally and internationally (Dauvergne, 2018). Policies dictate the amount of plastic that is allowed to be generated, the kinds of plastic that are produced, and the measures in place for their proper disposal. In 2021, global production of synthetic plastics reached 390.7 Mt per annum, joining a cumulative amount surpassing 8.3 billion metric tons—a mass sufficient to blanket the entire globe in cling film (Geyer, Jambeck, and Law, 2017; PlasticsEurope, 2022; Zalasiewicz, Waters, ... Yonan, 2016). It is estimated that, of the total sum of plastic, 30% is still in use, 10% has been recycled, and the remaining 60% has been disposed, either to landfills or to the natural environment (Geyer, Jambeck, and Law, 2017). Many plastics used today are exclusively single-use, and contribute to a hefty economic (80-120 billion USD annually) and environmental cost (“Strategy on Zero Plastic Waste,” 2018; Taniguchi, Yoshida, ... Oda, 2019). Consequently, plastic is being considered as a geological indicator for the Anthropocene era due to its widespread prevalence and endurance in global ecosystems (Zalasiewicz, Waters, ... Yonan, 2016).

Ubiquitous plastic usage has already resulted in environmental pollution, economic losses, and rising health concerns (Brahney, Hallerud, ... Hahnenberger, 2020; Rodrigues, Abrantes, ... Gonçalves, 2019; Shipton and Dauvergne, 2022). Roughly 80 Mt of plastic waste is improperly

disposed of each year, leading to accumulations of plastic in urban, natural, and oceanic areas (Lebreton and Andrady, 2019). Large plastic debris is known to cause external and internal injuries to marine animals and birds by entanglement or digestion (Derraik, 2002; Laist, 1997). Further, in urban settings, large plastic accumulates have resulted in concerns over increased stagnant water deposits, and damage to infrastructure (such as sewage blockages) (Shipton and Dauvergne, 2022). This danger is reflective of the materials durability, for while plastic is simple to make, it is resistant to degradation by photo-oxidation, hydrolysis, and/or thermal-oxidation (Chamas, Moon, ... Suh, 2020). It could be argued, additionally, that the uncontrolled, environmental degradation of plastics to smaller microplastics (1-1000 μm), or even nanoplastics (1 nm-1 μm) is not optimal. These small microplastics can spread through water and wind, thus allowing the absorption and broadcasting of many substances, including toxic compounds/additives and heavy metals (Rodrigues, Abrantes, ... Gonçalves, 2019; Wiesinger, Wang, and Hellweg, 2021; Wright and Kelly, 2017). These compounds have disproportionately severe affects on marine ecosystems, but are capable of causing irritation, organ impairment, carcinogenicity, fertility damage, and fatality to all living creatures with sufficient dosage (Rodrigues, Abrantes, ... Gonçalves, 2019). In recent years, research has shown that microplastics can bioaccumulate in mice tissues (particularly the liver, kidneys, and gut), and have been detected in healthy human blood (Deng, Zhang, ... Ren, 2017; Leslie, van Velzen, ... Lamoree, 2022). It is becoming clear by the increased evaluations of the environmental and toxicological fates of plastics, that more regulatory action is required to minimize the burden of the plastic life cycle.

1.2 Types of Plastics and Recycling Methods

Synthetic plastics are assembled by joining thousands of individual monomer units together by covalent linkages. The majority of these monomers are derived from naphtha, a product created from non-renewable fossil fuels, however bio based alternatives such as cellulose, starch etc. are gaining popularity (Leal Filho, Saari, ... Voronova, 2019). Fine tuning of plastic properties is possible by altering chemical composition, degree of branching, branch lengths, polarities, sheet orientations, and crystallinity (Baker and Mead, 2000). Additionally, additives such as colorants, antistatic agents, fillers etc., can further modify product characteristics (Wiesinger, Wang, and Hellweg, 2021). Unfortunately, this simple building block process is not easily reversible, making the recovery and recycling of plastics challenging. Further, due to the need for the non-trivial separation, and sorting of polymer waste by chemical composition and color, the diversification of products, while industrially valuable, has likely resulted in complications to the plastic recovery process (da Silva and Wiebeck, 2020).

Recycling alone is currently not a sustainable solution to the plastic burden—rather it is meant to support progress in reducing virgin plastic need. The most economical form of plastic restoration is by mechanical or physical recycling (Figure 1.1) (Maisels, Hiller, and Simon, 2022). During this process, the underlying chemical structure of plastics is not altered; instead, polymers are ground, melted, and reshaped into sheets or plastic pellets for reuse (da Silva and Wiebeck, 2020; Selke, 2000). Typically, due to poor consumer separation and contaminants (such as dyes, particulates), obtaining high quality and homogenous plastics with this method is intractable (da Silva and Wiebeck, 2020). Comparatively, thermo-chemical recycling results in a fundamental change to the underlying plastic polymer, and is completed by hydrolysis, solvolysis, glycolysis, hydrocracking, catalytic cracking, pyrolysis, and gasification (Ragaert, Delva, and Van Geem,

2017). Condensation plastics—such as PET, nylon, or polyurethane—are especially amendable to chemical depolymerization (Selke, 2000). Most methods rely on high heat in order to decompose large feedstock molecules into a complex mixture of small products (Idumah, 2022; Selke, 2000). Here, the product distribution is dependent on the feedstock composition, temperatures, pressures, catalyst presence, additive presence, and atmospheric conditions (Ragaert, Delva, and Van Geem, 2017; Zeller, Netsch, ... Stapf, 2021). If reasonably pure materials are recovered, they are able to be re-used for plastic production, however it is more typical that the resulting products are utilized as secondary fuels (Idumah, 2022; Selke, 2000). Akin to thermal recycling methods is incineration—though as no usable products are being formed, it is considered a lower level of recycling—in which plastic waste is burned to produce carbon dioxide (g), water, and energy (steam, heat, electricity) (Selke, 2000). Notably, while all high temperature routes have promise in the recovery of valuable products/energy and do not require thorough sorting of plastic feedstocks, they are carefully regulated due to opportunities for gas pollution and hazardous by-product formation (Idumah, 2022; Selke, 2000). Thus, it becomes clear that while many methods for recycling exist (Figure 1.1), they are prone to creating substandard plastics, pollution, and require large amounts of chemicals or heat.

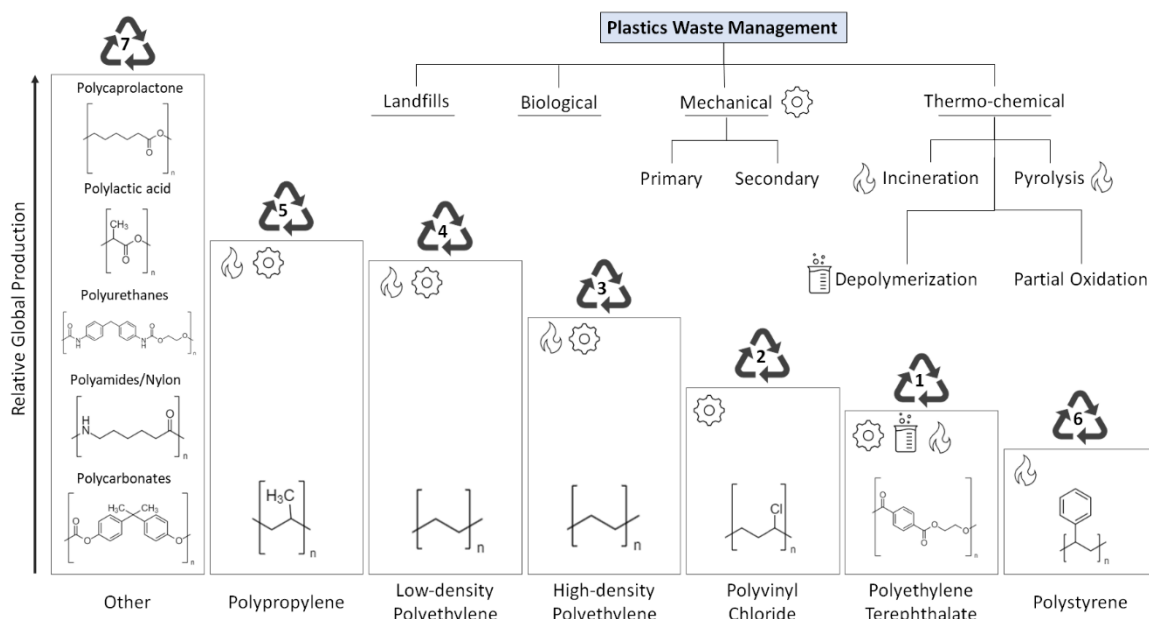


Figure 1.1 Relative production of plastics by type and recycling options. Plastics are generally identified as one of seven types: polypropylene, low-density polyethylene, high density polyethylene, polyvinyl chloride, polyethylene terephthalate, polystyrene and other. Relative amounts of global production are shown by height of boxes. Main recycling options include pyrolysis (flame icon), mechanical remodelling (gear icon), and depolymerization (beaker icon). Note that plastics in the ‘other’ category (7) can undergo different recycling methodologies based on their properties, and are thus not collectively defined. Though microbial degradation has been seen for many plastic types, the efficiencies are not necessarily adequate for large scale recycling. The waste management flowchart was adapted from (Idumah, 2022). The plastic distribution chart was adapted from (Lear, Maday, ... Pantos, 2022).

The ideal goal for the plastics industry is the creation of a circular economy in which virgin plastic production is limited, and the sustainable resurrection of high-quality plastic from waste is possible. One recycling method that has caught considerable interest in recent years is microbial or enzymatic plastic degradation. Distinctly, this method is possible at lower temperatures, in aqueous solution, and can deliver high quality monomers for reuse (Tamoor, Samak, ... Xing, 2021). Many microbial organisms have been reported to possess plastic degrading abilities (Figure 1.2), and some are capable of using plastics as a major carbon source (Mohanam, Montazer, ...

Levin, 2020; Priya, Dutta, and Daverey, 2022; Skariyachan, Setlur, ... Vasist, 2017). The industrial leveraging of many of these degradation pathways are not yet possible largely due to insufficient degradation rates. The structural resilience of plastics is the main limitation: some types lack reactive functional groups (polyethylene, polypropylene), structures can be highly crystalline, surfaces are often naturally hydrophobic, and some plastics feature antibacterial bioetching (Idumah, 2022; Taniguchi, Yoshida, ... Oda, 2019; Tripathy, Sen, ... Briscoe, 2017). Relevantly, there is a lack of natural enzymes capable of directly cleaving carbon-carbon bonds. Accordingly, enzymatic degradation of polyethylene etc. necessitates carbonyl formation by abiotic activities such as mechanical comminution, temperature or pH fluctuations, ultraviolet irradiation and oxidation (Ghatge, Yang, ... Hur, 2020; Lin, Jin, ... Fei, 2022; Taniguchi, Yoshida, ... Oda, 2019). Therefore, there is rightly a large amount of criticism for the enzymatic recycling of more inert plastics (Lear, Maday, ... Pantos, 2022). In comparison, plastics with functional groups, such as PET and nylon, are typically more susceptible to enzymatic attack. Here, an important constraint is rigid polymer chains and high levels of crystallinity, which cloak and protect the bulk plastic structure (Taniguchi, Yoshida, ... Oda, 2019). Ergo, to obtain favourable degradation rates, it is necessary to increase enzyme access to the bulk plastic, which is improved by increasing reaction temperatures to above the plastics glass transition range (T_g) (Kawai, Kawabata, and Oda, 2019). In this range, crystalline plastics become more amorphous: hydrogen bonds are weakened, chain flexibility is increased, and chain orientations/packing is randomized (Baker and Mead, 2000; Kawai, Kawabata, and Oda, 2019). The critical goal is then to engineer the enzymes and pathways involved to withstand increased reaction temperatures, and to holster high degradation rates for industrial upscaling.

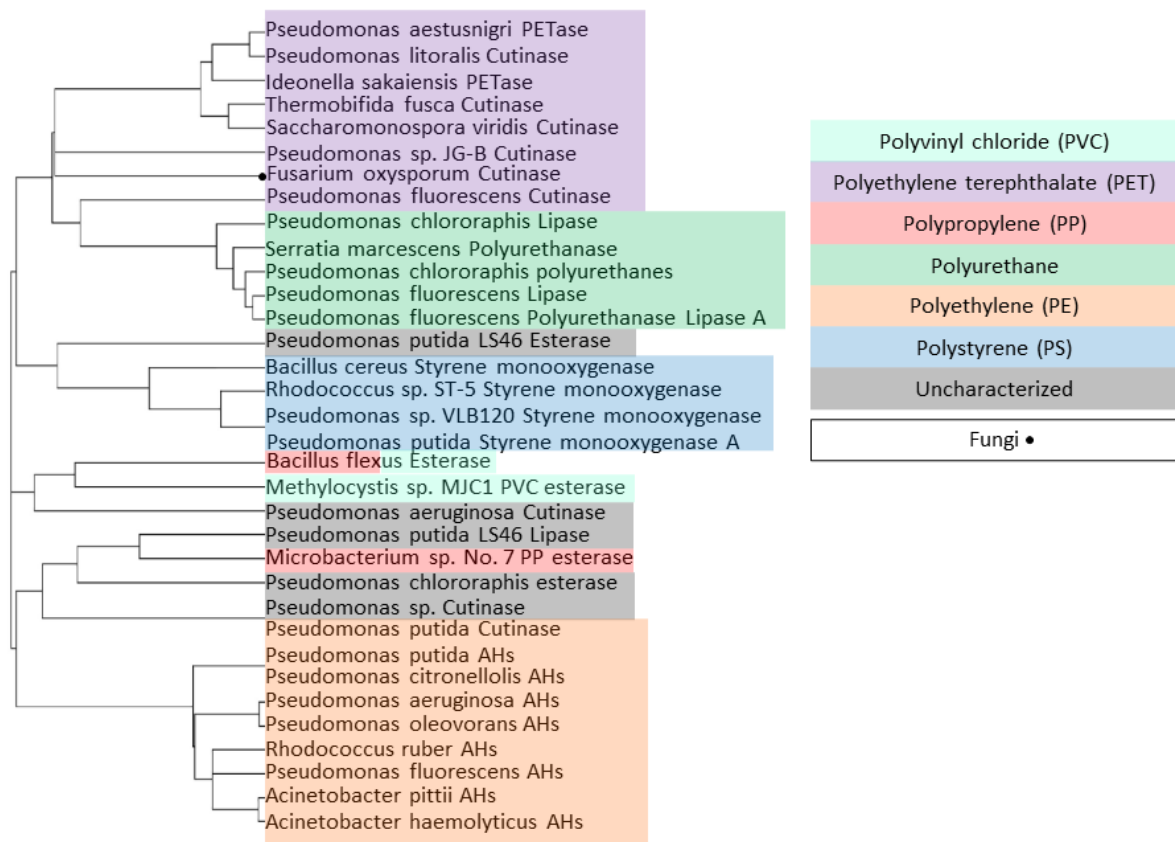


Figure 1.2 Phylogenetic relationships between natural plastic degrading enzymes. The above phylogenetic tree was created by (Mohan, Montazer, ... Levin, 2020), and is based on the amino acid sequence homologies of plastic degrading enzymes. The main plastic type degraded by phylogenetic group is shown by coloured boxes. It should be noted that this list is not comprehensive, as natural enzymes in bacterial, fungal, and archaea taxa have also shown plastic degrading abilities (Ghatge, Yang, ... Hur, 2020; Priya, Dutta, and Daverey, 2022).

1.3 Polyethylene Terephthalate (PET) Degrading Enzymes

1.3.1 Natural PET Hydrolases

Amidst the PET polymer structure are ester linkages—a functional group that is hydrolysable by a horde of enzymes (typically alpha-beta hydrolases which include lipases, esterases, and cutinases). This superfamily of proteins, described well in the ESTHER database, can vary greatly in their stability and substrate specificity (Lenfant, Hotelier, ... Chatonnet, 2012). Many enzymes within this super-family have demonstrated substrate ambiguity towards PET (Table 1)—specifically cutinases which can act upon a range of substrates from insoluble polymers such as cutin to soluble fatty acid esters, and triglycerides. Notably, several of these enzymes belong to organisms found growing in crowded plastic environments. One such organism, *Ideonella sakaiensis*, has proven ability to employ PET as a main source of carbon—suggesting the existence of a metabolic linkage between PET products and central carbon intermediates (Figure 1.3) (Yoshida, Hiraga, ... Oda, 2016). Indeed, the enzyme responsible for its ability to degrade plastic, Is. PETase, possesses one of the most efficient natural PET degradation rates at ambient temperatures (Yoshida, Hiraga, ... Oda, 2016). Comparably, PET degrading enzymes with higher stability, such as LCC ($T_m \sim 85^\circ\text{C}$, vs Is. PETase $T_m \sim 42^\circ\text{C}$), are capable of a higher total degradation level of PET due to increased substrate pliancy at elevated temperatures (Guo, Vanga, ... Syrén, 2022; Tournier, Topham, ... Marty, 2020). The ideal temperature of activity for PET recycling may be dependent on the application process (i.e., use of a bioreactor with living organisms, or extracellular enzyme production/application), however, the ensured stability, longevity, solubility, and activity of the enzyme remains a valuable objective for industrial application. Currently, all known PET hydrolyzing enzymes are considered to operate in a non-specific manner, and they do not have the efficiencies required for direct recycling use.

Table 1. Characterized Natural PET Hydrolysing Enzymes. Known enzymes with PET degrading abilities are listed below, and are referenced by their GenBank accession number and PDB structure (if available). Temperature range describes range in which enzyme activity against PET substrates was reported—this does not necessarily correlate to a temperature optimum. Based on (Kawai, Kawabata, and Oda, 2019; Taniguchi, Yoshida, ... Oda, 2019). ^aPDB code reported for enzymes are representative, and may be one of many solved. ^bBiodegradable PET contain additional aliphatic-aromatic co-polyesters and are more easily digestible by esterases.

Enzyme	GenBank	PDB ^a	Organism	Temp Range (°C)	Substrate	Reference
PET 27	WP_111881932		<i>Aequorivitia</i>	30	PET foil	(Zhang, Perez-Garcia, ... Streit, 2022)
BsEstB	ADH43200.1		<i>Bacillus subtilis</i> 4P3-11	40-45	3-PET	(Ribitsch, Heumann, ... Guebitz, 2011)
PET-C2-7			<i>Bacillus toyonensis</i>	37	BHET on agar plate	(Qiao, Hu, ... Du, 2022)
PET-C2-8			<i>Bacillus cereus</i>	37	BHET on agar plate	(Qiao, Hu, ... Du, 2022)
Bhr-PETase	GBD22443.1	7EOA	<i>Bacterium HR29</i>	60-80	PET powder	(Xi, Ni, ... Qian, 2021)
Bb PETase		7CWQ	<i>Burkholderiales</i> bacterium	30-40	Low crystallinity PET film	(Sagong, Son, ... Kim, 2021)
Cbotu_Est A	A5I055	5AH1	<i>Clostridium botulinum</i> ATCC3502	50	PET film	(Biundo, Reich, ... Guebitz, 2018)
FsC		1CEX	<i>Fusarium solani</i> pisi	30-60	PET film	(Yoshida, Hiraga, ... Oda, 2016)
HiC		4OYY	<i>Humicola insolens</i>	40-85	3-PET	(Ronkvist, Xie, ... Gross, 2009)
Is PETase	GAP38373.1	6EQE	<i>Ideonella sakaiensis</i> 201-F6	20-45	PET film	(Yoshida, Hiraga, ... Oda, 2016)
PET 30	KIA88187.1	7PZJ	<i>Kaistella jeonii</i>	30	PET foil	(Zhang, Perez-Garcia, ... Streit, 2022)
Un-5			<i>Kineococcus endophyticus</i>	37	BHET on agar plate	(Qiao, Hu, ... Du, 2022)
PET-C2-A			<i>Lysinibacillus fusiformis</i>	37	BHET on agar plate	(Qiao, Hu, ... Du, 2022)
Ple628	OK558825	7VMD	Marine Microbial Consortium	30	PET nanoparticles	(Meyer Cifuentes, Wu, ... Öztürk, 2022)
Ple629	OK558824	7VPA	Marine Microbial Consortium	30	PET nanoparticles	(Meyer Cifuentes, Wu, ... Öztürk, 2022)

PSH1		7W6O	Metagenome Derived	72	PET film, nanoparticles	(Pfaff, Gao, ... Wei, 2022)
PHL-7, PES-H1		7NEI	Metagenome Derived	65-75	Low crystallinity PET film	(Sonnendecker, Oeser, ... Zimmermann, 2022)
PES-H2			Metagenome Derived			(Zimmermann, WEI, ... SCHMIDT, 2019)
Un-10			<i>Niveispirillum cyanobacteriorum</i>	37	BHET on agar plate	(Qiao, Hu, ... Du, 2022)
Tha_Cut I	ADV92525.1		<i>Thermobifida alba</i> DSM43185	50	3-PET	(Ribitsch, Acero, ... Guebitz, 2012)
Est 119	BAK48590.1	3VIS	<i>Thermobifida alba</i> AHK119	45-55	Biodegradable PET ^b	(Hu, Thumarat, ... Kawai, 2010)
Thc_Cut I	ADV92526.1		<i>Thermobifida cellulositytica</i> DSM44535	50	3-PET, PET film	(Herrero Acero, Ribitsch, ... Guebitz, 2011)
Thc_Cut II	ADV92527.1	5LUJ	<i>Thermobifida cellulositytica</i> DSM44535	50	3-PET, PET film	(Herrero Acero, Ribitsch, ... Guebitz, 2011)
Tcur0390	CDN67546.1		<i>Thermomonospora curvata</i> DSM 43183	50	PET nanoparticles	(Wei, Oeser, ... Zimmermann, 2014)
Tcur1278	CDN67545.1		<i>Thermomonospora curvata</i> DSM 43183	50-60	PET nanoparticles	(Wei, Oeser, ... Zimmermann, 2014)
Tf42_Cut I	ADV92528.1		<i>Thermobifida fusca</i> DSM44342	50	3-PET, PET film	(Herrero Acero, Ribitsch, ... Guebitz, 2011)
TfH	WP_011291330.1		<i>Thermobifida fusca</i> DSM43793	30-60	PET film	(Yoshida, Hiraga, ... Oda, 2016)
Tfu_0883	Q47RJ6		<i>Thermobifida fusca</i> YX	60	PET fabric	(Silva, Da, ... Cavaco-Paulo, 2011)
Tf. Cut II	CBY05530.1	4CG1	<i>Thermobifida fusca</i> KW3	65-75	PET film	(Furukawa, Kawakami, ... Miyamoto, 2019)
Thh_Est	AFA45122.1		<i>Thermobifida halotolerans</i> DSM44931	50	PET film	(Ribitsch, Acero, ... Guebitz, 2012)
Rg PETase		7DZT	<i>Rhizobacter gummiphilus</i>	30-40	PET film	(Sagong, Son, ... Kim, 2021)
Cut 190 S226P/R228S	BAO42836.1 (WT)	4WFI	<i>Saccharomonospora viridis</i> AHK190	60-65	PET film	(Kawai, Oda, ... Tanokura, 2014)
Un-5-5			<i>Shewanella xiamenensis</i>	37	BHET on agar plate	(Qiao, Hu, ... Du, 2022)

Un-14			<i>Sinirhodobacter hankyongi</i>	37	BHET on agar plate	(Qiao, Hu, ... Du, 2022)
Un-C2-8			<i>Staphylococcus epidermidis</i>	37	BHET on agar plate	(Qiao, Hu, ... Du, 2022)
PET5	R4YKL9		<i>Oleispira antarctica</i> RB-8	50	PET nanoparticles (agar)	(Danso, Schmeisser, ... Streit, 2018)
PE-H		6SBN	<i>Pseudomonas aestusnigri</i>	30	Low crystallinity PET film	(Bollinger, Thies, ... Jaeger, 2020)
PmC			<i>Pseudomonas mendocina</i>	50	Low crystallinity PET film	(Ronkvist, Xie, ... Gross, 2009)
Un-3-2			<i>Pseudomonas protegens</i>	37	BHET on agar plate	(Qiao, Hu, ... Du, 2022)
PET12	AKJ29164.1		<i>Schlegelella brevitalea</i>	50	PET nanoparticles (agar)	(Danso, Schmeisser, ... Streit, 2018)
LCC	AEV21261.1	4EB0	Uncultured bacterium	30-80	PET film	(Yoshida, Hiraga, ... Oda, 2016)
PET2	ACC95208		Uncultured bacterium	50	PET nanoparticles (agar)	(Danso, Schmeisser, ... Streit, 2018)
PET6	SHF85073.1	7Z6B	<i>Vibrio gazogenes</i>	50	PET nanoparticles (agar)	(Danso, Schmeisser, ... Streit, 2018)

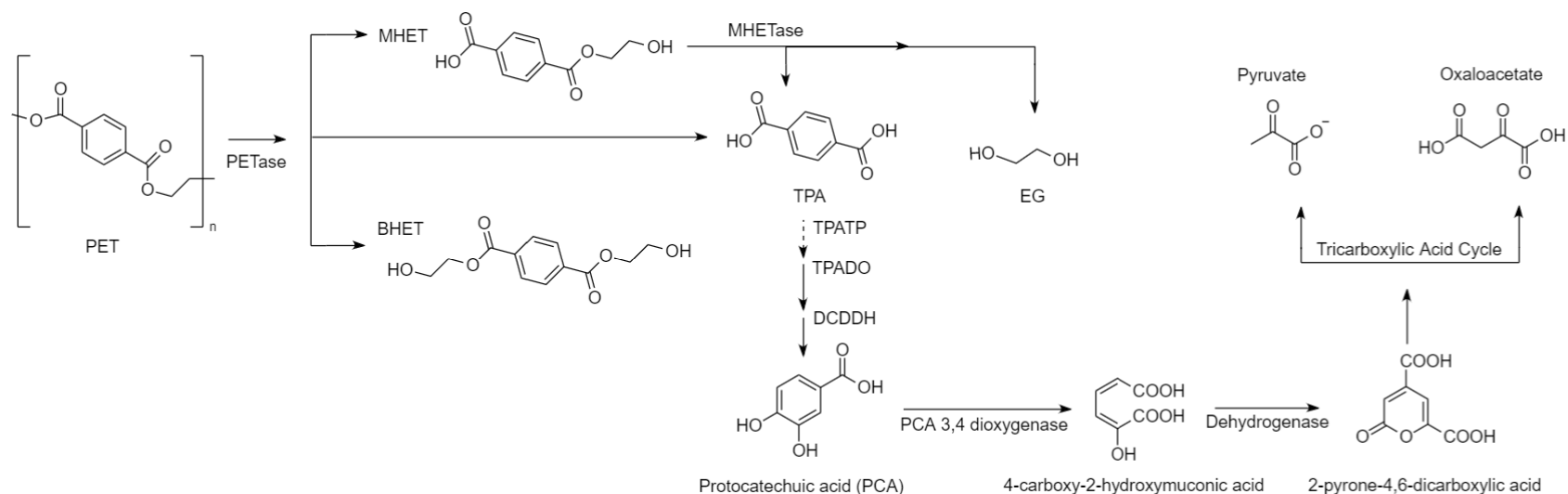


Figure 1.3 Suggested catabolic pathway for PET degradation to oxaloacetate and pyruvate. PET plastic degradation pathway was proposed by (Priya, Dutta, and Daverey, 2022; Yoshida, Hiraga, ... Oda, 2016), and is based on analogies to degradation systems confirmed in other bacteria. Here, extracellularly expressed PETase degrades PET into MHET (major product), BHET, or TPA. An extracellular lipoprotein MHETase further assists in degrading the major product MHET into a transportable molecule, TPA. TPA is imported through the TPA transporter (TPATP) (Hosaka, Kamimura, ... Masai, 2013), and then subsequently broken down into protocatechuic acid (PCA) by TPA 1, 2-dioxygenase (TPADO), and 1,2-dihydroxy-3,5-cyclohexadiene-1,4-dicarboxylate dehydrogenase (DCDDH). Breakage of the PCA aromatic ring is completed by PCA 3,4-dioxygenase, and the ensuing 4-carboxy-2-hydroxymuconic acid is converted to 2-pyrone-4,6-dicarboxylic acid. Finally, 2-pyrone-4,6-dicarboxylic acid can be degraded to 4-oxalomesaconate (as seen in lignin degradation), which is further transformed into TCA intermediates pyruvate and oxaloacetate (Perez, Kontur, ... Noguera, 2019). This figure was created using ChemDrawDirect (PerkinElmer).

Intriguingly, crystal structures for all discovered PET hydrolases (excluding PDB structures for Cbotu_Est A, HiC, FsC) reveal a canonical α/β fold composed of 8-9 beta sheets surrounded by 6-7 alpha helices (Figure 1.4A). They each possess an exposed catalytic site formed by the triad Ser-His-Asp and a stabilizing oxyanion hole (Figure 1.4A). Unlike similarly structured lipases, which are required to act at lipid-water interfaces, PET hydrolases typically lack protective lids and have a shallower, more open active site (Kawai, Kawabata, and Oda, 2019). The catalytic mechanism, which involves the nucleophilic attack of the substrate ester bond by Ser, the formation of a covalent tetrahedral intermediate, and the release of the product via hydrolysis is conserved amongst this enzyme class (Figure 1.4B). Due to the nature of the superficial active site, and large polymer structure of the PET substrate, the interaction between PET and the enzyme surface is crucial for activity—indeed activity and absorption of the enzyme to PET film surfaces seem to be correlated (Furukawa, Kawakami, ... Miyamoto, 2018; Kozłowski, 2016; Sagong, Son, ... Kim, 2021). Thus, it is especially interesting that known PET hydrolyzing enzymes have a range of predicted pI values (4-9.5), suggesting they may have different interaction efficiencies with the PET film, varying approach orientations, and diverse optimal pHs (Figure 1.4A). To date, the co-crystallization of a PET hydrolase with a PET polymer has not yet been possible, and only PDB 5XH3 contains an illustration of a single PET moiety bound to Is. PETase (Han, Liu, ... Guo, 2017). Thus, docking has been extensively used to study the substrate-enzyme relationship, particularly in consideration of the conformations possible for the PET substrate (i.e. *gauche* or *trans*) (Charupanit, Tipmanee, ... Limsakul, 2022; Fecker, Galaz-Davison, ... Ramírez-Sarmiento, 2018; Guo, Vanga, ... Syrén, 2022, p. 20; Joo, Cho, ... Kim, 2018; Wei, Song, ... Zimmermann, 2019). In addition to the above, all PET hydrolases possess a conserved C-terminal disulfide, which contributes to the thermodynamic and kinetic stability of the enzyme (Sulaiman, You, ...

Kanaya, 2014). Thus, while many overall characteristics between PET hydrolases are shared, it is prudent to focus on more subtle variations in their sequences/structures that lead to known differences in their activity and stability.

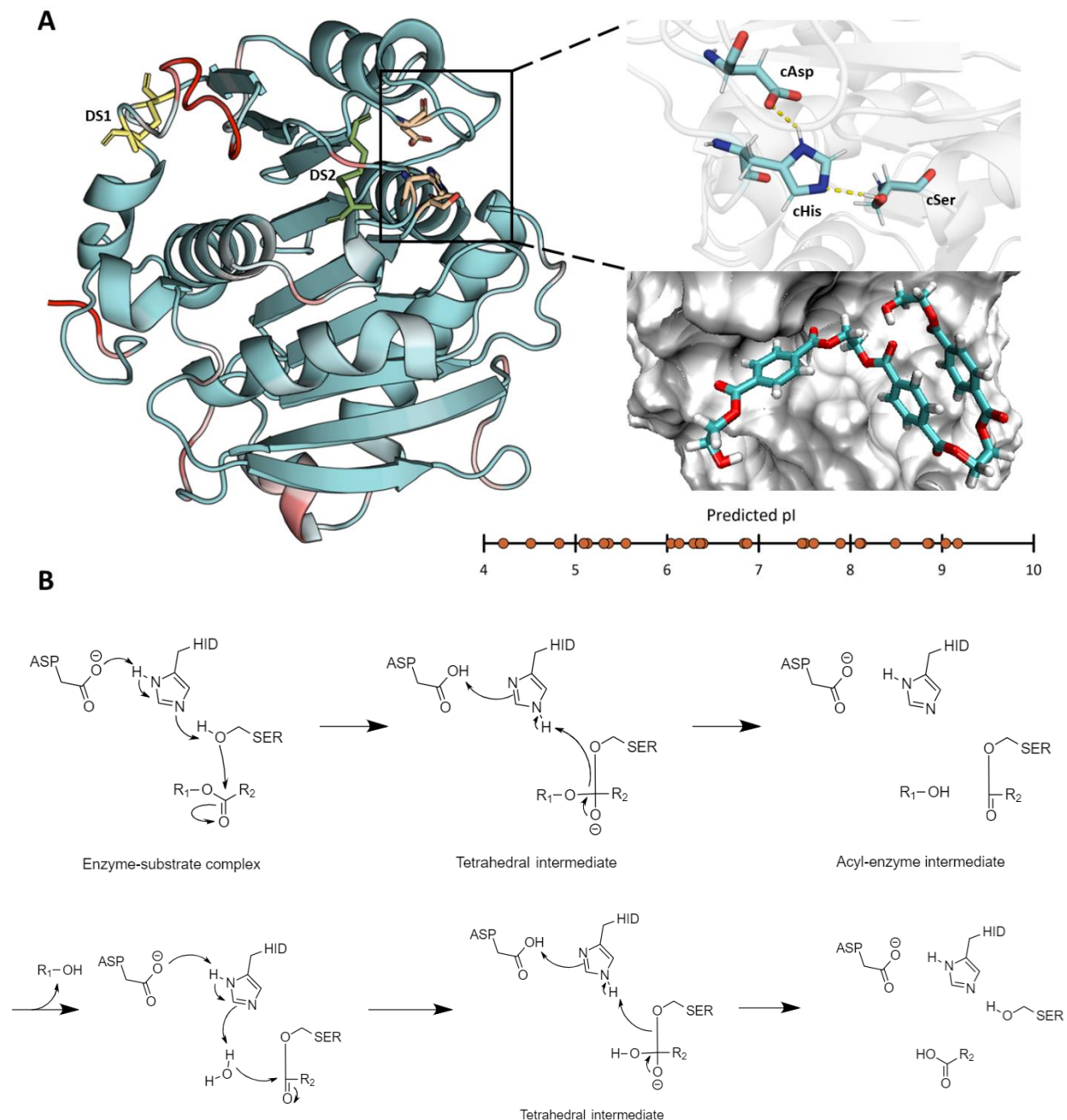


Figure 1.4 Structural characteristics of PET hydrolases. **A** Structure of Tf. Cut II (4CG1) is shown on the left. It is colored by structural similarity to other PET hydrolases, such that diverse structural areas are colored red to white, while more conserved areas are cyan. The catalytic triad is shown in peach, and is within the focus box on the top right. The conserved disulfide bond is shown as sticks, yellow (DS1). Is. PETase has an additional disulfide bond near the active site—the location is displayed as sticks, green (DS2). An example docking pose of 3-PET and Tf. Cut II generated using Schrodinger by induced fit docking is also shown on the right. It illustrates the surface binding nature of the PET molecule and the tendency of the PET polymer to self-associate. A small 2D plot of the predicted pI of PET hydrolases is

presented to demonstrate the possible diverse nature of the enzyme interaction surface. **B** Classic catalytic mechanism for serine proteases. The nucleophilic serine is activated by a charge-relay network in Asp-His, resulting in the attack of the substrate ester bond and the formation of the subsequent tetrahedral and covalent acyl-enzyme intermediates. In the second part of the reaction, a water molecule attacks the Ser-substrate complex, forming another tetrahedral intermediate, which finally leads to the release of the product and the regeneration of the catalytic triad. This figure was created using ChemDrawDirect (PerkinElmer).

1.3.2 Engineered PET Hydrolases

In a short time of seven years since the discovery of Is. PETase in 2016, there has been a tremendous effort to engineer highly active and stable PET hydrolyzing enzymes (Appendix A). Engineering methods are diverse and have included a myriad of engineering approaches—targeted alanine scanning, surface-charge conversion, docking or structural based design, consensus matching or homolog matching, addition of disulfide bonds, immobilization, fusion protein generation, and directed evolution (Brott, Pfaff, ... Bornscheuer, 2022). In addition, based on the vast database created for single mutations, multi-mutant engineering has been actively explored. Notably, some approaches use unique computational methods; these are based on mutation to less common residues in PET hydrolyzing homologs (Premuse), machine learning (GRAPE, MutCompute), and available stability predictors (PROSS) (Cui, Chen, ... Wu, 2021; Lu, Diaz, ... Alper, 2022; Meng, Yang, ... Tian, 2021; Rennison, Winther, and Varrone, 2021). Among the most accomplished thermostable variants created by any approach are Tf. Cut II/D204C-E253C/D174R/G205D ($T_m = 94.6\text{ }^\circ\text{C}$), DuraPETase/S23E/N204K ($T_m = 83.4^\circ\text{C}$), and LCC/D203C-S248C/F208W/N211M ($T_m = 98.1\text{ }^\circ\text{C}$) (Lu, Diaz, ... Alper, 2022; Then, Wei, ... Zimmermann, 2016; Tournier, Topham, ... Marty, 2020). As their overall stabilities were improved, and they typically showed high temperature optimums for activity, they are promising variants for the degradation of plastics at high temperatures. Alternatively, FastPETase and

ThermoPETase have notably stronger PET degradation activity at mild temperatures (30-50 °C) making them more amendable for bioreactor based recycling (Lu, Diaz, ... Alper, 2022; Son, Cho, ... Kim, 2019). In all of the above cases the creation of novel PET hydrolyzing variants with high potential for enzymatic recycling was possible by the iterative combination and testing of several stabilizing/activating point mutations. This raises the question of whether Ancestral Sequence Reconstruction (ASR), which is known for its one-shot, semi-rational design strategy, could produce a multi-mutant with similar characteristics. For the remainder of this thesis ASR will be critically assessed as a strategy for the design of novel PET hydrolysing enzymes.

1.4 Ancestral Sequence Reconstruction as a Method for Creating Promiscuous and Stable Proteins

1.4.1 Introduction to Ancestral Sequence Reconstruction

In 1963, Linus Pauling and Emile Zuckerkandl having noticed a persistent amino acid pattern among hemoglobin polypeptides, first suggested the possibility of a shared common ancestor (Pauling, Zuckerkandl, ... Lövstad, 1963). This idea, combined with the ascent of computational prowess and more scientific interest, has resulted in the development of a technique which allow properties and histories of these shared ancestors to be probed. Thus, ancestral sequence reconstruction (ASR) has been used to examine the emergence of enzyme catalysis (especially from inactive scaffolds), and specificity (Bar-Rogovsky, Hugematter, and Tawfik, 2013; Clifton, Kaczmariski, ... Jackson, 2018; Hultqvist, Åberg, ... Jemth, 2017). Remarkably, reconstructed ancestors have often displayed characteristics that are desirable objectives for protein engineers—thermostability, substrate promiscuity, increased activity, and increased solubility—rendering it a promising semi-rational design approach (Barruetaña, Alonso-Lerma, ... Perez-Jimenez, 2019; Gumulya, Baek, ... Gillam, 2018; Whitfield, Zhang, ... Jackson, 2015). Indeed, thermostability increases ranging up to 30 °C as well as improved pH tolerance from extant enzymes have been observed (Babkova, Sebestova, ... Damborsky, 2017; Barruetaña, Alonso-Lerma, ... Perez-Jimenez, 2019; Gumulya, Baek, ... Gillam, 2018; Risso, Gavira, ... Sanchez-Ruiz, 2013; Whitfield, Zhang, ... Jackson, 2015; Zakas, Brown, ... Doering, 2017). This phenomenon may be due to the statistical likelihood of ASR to favour consensus residues, where it is argued that an enduring amino acid contributes more to the average stability than infrequent amino acids (Porebski and Buckle, 2016). However, in one study, in which consensus failed to produce an active variant, an ancestor with increased promiscuity and thermostability (compared

to consensus and wild-type) was obtained (Hendrikse, Charpentier, ... Syrén, 2018). Thus, this suggests that ASR may be more likely to consider evolutionary context, such as coevolution within residue networks (Hendrikse, Charpentier, ... Syrén, 2018). Many reconstructed ancestors have higher and broader activity, and can be used as starting points for optimization of enzymes with novel functions (Babkova, Sebestova, ... Damborsky, 2017; Barruetabeña, Alonso-Lerma, ... Perez-Jimenez, 2019; Gumulya, Baek, ... Gillam, 2018; Hendrikse, Charpentier, ... Syrén, 2018; Nakano, Minamino, ... Ito, 2019; Wilding, Peat, ... Jermin, 2017; Zakas, Brown, ... Doering, 2017). The potential of ASR to navigate the stability-function trade-off—perhaps by alternating the conformational flexibility of key regions—is intriguing, and may further reveal previously unpredicted sites for targeted engineering (Gamiz-Arco, Gutierrez-Rus, ... Sanchez-Ruiz, 2021; Tokuriki, Stricher, ... Tawfik, 2008).

1.4.2 Maximum likelihood for Ancestral Sequence Reconstruction

Three statistical methods have been applied in the determination of ancestral states: parsimony (Sankoff, 1975), maximum likelihood (Schluter, Price, ... Ludwig, 1997), and the Bayesian approach (Pagel, Meade, and Barker, 2004). The most commonly used approach is maximum likelihood, perhaps because it is more likely to retain the most probable residue for each ancestor state (Groussin, Daubin, ... Tannier, 2016). Very briefly, as an example, the likelihood of the tree in Figure 1.5—or the probability of observing the sequences S1-3 and A1-2—will be computed (Equation 1). Assuming that all amino acid sites (or columns in the multiple sequence alignment) mutate independently, then the total likelihood is the product of the likelihood values at individual amino acid sites. These individual likelihoods can be explained by simple probability

terms (Pr) based on the Markov property—which states that the probability of a sequence can be determined by the sequence of its parent node and the length of the branch between them.

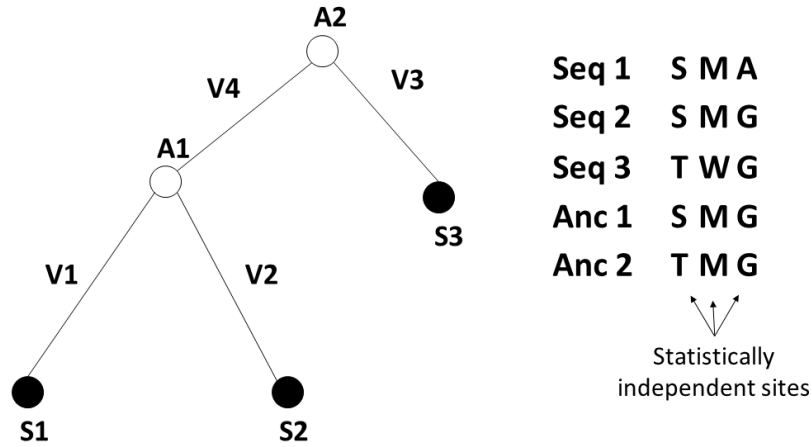


Figure 1.5 Example of Phylogenetic Tree to demonstrate main ideas in ASR. Phylogeny of three extant proteins (labelled S1-S3) with internal nodes (A1, A2) representing ancestral sequences. Example sequences are given for proteins on the right. The length of the vertices is represented by V1-V3. For maximum likelihood, columns in the MSA are considered statistically independent.

$$\begin{aligned} \text{Likelihood of current observations at site 1} &= \Pr(A2: T) * \Pr(S3: T | A2: T, V3) * \\ &\Pr(A1: S | A2: T, V4) * \Pr(S2: S | A1: S, V2) * \Pr(S1: S | A1: S, V1) \end{aligned} \quad (1)$$

Here, conditional probability terms are computed based on the chosen evolutionary model. In the scenario of ancestral sequence reconstruction, the sequence of the ancestors is not known, and instead sequences are proposed such that the likelihood of the tree is maximized. This is a computationally consuming problem that scales with increasing tree size, as calculation of the full likelihood would require comparisons of every possible ancestor sequence at every ancestral node: even the small tree shown in Figure 1.5 would require $20^2 * 3$ separate likelihood calculations (Groussin, Daubin, ... Tannier, 2016; Pupko, Pe, ... Graur, 2000). Accordingly, dynamic programming which isolates smaller sub-trees, and calculates the likelihood thereof, is capable of

scaling the problem into linear terms, thereby reducing computation difficulty (Groussin, Daubin, ... Tannier, 2016; Pupko, Pe, ... Graur, 2000). Importantly, as is shown, the results from ASR are highly dependent on the accuracy of the model parameters: notably, the alignment quality, the optimized tree topology/branch lengths, and evolutionary model.

Given the extent of available sequence homologs in the PET hydrolase family, it was alluring to resurrect its ancestral enzymes, with the primary goal that they would indeed bolster increased thermostability and promiscuity. As well, there was an interest in the types and number of mutations that would be predicted, and the mechanism in which they would infer changes to the overall structure—stable and active reconstructed proteins with 100 site mutations have been seen (Risso, Gavira, ... Sanchez-Ruiz, 2013)! To this effect, maximum likelihood ASR is used to predict the sequence for two ancestral nodes of interest. These were expressed *in vitro*, purified, and investigated for altered characteristics compared to the wild-type Tf. Cut II. In addition, the observed results are examined computationally in order to understand how the combination of multiple mutations may have affected overall protein structure.

2 Materials and Methods

2.1 In Silico

2.1.1 Generation of Ancestor Sequences

Sequences were obtained with basic local alignment search tool (BLASTp) using target sequences Tf. Cut II or Is. PETase. A number of databases were explored using default search parameters: Refseq, non-redundant, reference protein, model organisms, Uniprot/SwissProt, patent division, PDB, WGS metagenomics, transcriptome shotgun assembly (short queries = ‘yes’, expect threshold = 0.05, word size = 6, max matches in a query range = 0, Matrix = ‘BLOSUM62’, gap costs = existence 11 - extension 1, compositional adjustments = ‘conditional compositional score matrix adjustment’). The majority of hits were in the non-redundant database, which is composed of several sets of protein databases (Genbank, PDB, SwissProt, PIR, PRF). Initial filtering of all sequences was completed based on E-values (cut-off of 1E-20 used), coverage (>75% required), and % identity (>30% required). At this stage, a total of 4933 sequences remained (Table 2). Finally, analogous sequences were identified and removed by GENE ID or by >95% sequence identity, such that only one of them remained as representation.

Table 2 BLASTp hits for Tf. Cut II

Database	Total # Sequences	# Hypothetical sequences	Sequences Remaining Post-Initial Filtration
Refseq	1372	67	966
Non-redundant	4620	680	2792
Reference Protein	1113	4	1108
Model Organisms	11	0	1
Uniprot/Swissprot	10	0	3
Patent Division	33	0	15
PDB	43	0	43
WGS Metagenomics	137	84	4
Transcriptome Shotgun Assembly	40	2	1
Total	7379	837	4933

Filtering became difficult at this stage, as 888 candidate sequences remained, and there was no ideal approach to narrow the number of sequences (as explained in section 3.1). Allocation of sequences to groups by their functional annotation was proposed as a method for separation, especially as the cutinase sub-class was being recognized for their promiscuous activity towards PET (Table 3). Hence, sequences chosen for alignment construction were tagged as cutinases [24], PETase [1], thermo [25], lipase [25], poly (ethylene terephthalate) [1], and polyester [1]. This assortment includes a few unique sequences that were labelled as being plastic-related, and a small subset of the lipase group (closest sequence identities to Tf. Cut II). The argument for the inclusion of close lipase sequences, rather than some other class, was to potentially bias the ancestor sequence to prefer residues that bear potential hydrophobic substrate binding. Similarly, in a second alignment, a set of 25 of the closest PET homologs were included. After this step, duplicates were removed, as it was possible for enzymes to belong to two annotation groups. First (APPENDIX B) and second (APPENDIX C) sequence sets are respectively named “Cut-PETase” and “Cut-Therm”, in correspondence to the ancestor reconstructed from each.

Table 3 Annotation based separation of remaining BLASTp candidates

Annotation	# of Sequences
Cutinase	24
Alpha/Beta Hydrolase	299
Thermo	25
PETase	1
Thermobifida	6
Triacylglycerol Lipase	27
Lipase	163
Acetyl hydrolase	3
Esterase	7
Peptidase	1
Dienelactone Hydrolase	297
Polyester	1
Poly (ethylene terephthalate)	1
Phospho	11
Lectin	17
T9ss	26
Discoidin	4
Cellulose	9
Other	23

MSA Cut-PETase (61 Sequences)	Cutinase [24] PETase [1] Thermo [25] Lipase [25] Poly (ethylene terephthalate) [1] Polyester [1]
----------------------------------	---

MSA Cut-Therm (75 Sequences)	Cutinase [24] PETase [1] Thermo [25] Lipase [25] Poly (ethylene terephthalate) [1] Polyester [1] Closest PETase Homologs [25]
---------------------------------	---

Following choice of sequences, alignments were created using MAFFT, and checked manually for proper alignment (Kato, 2002). Sequences that displayed poor fits were removed, specifically if they caused extra insertions between conserved sections. Both alignments were checked/edited using Guidance and obtained final scores of 0.985763 and 0.979989 (Cut-PETase and Cut-Therm respectively) (Penn, Privman, ... Pupko, 2010). Phylogenetic trees were created

using IQTree (Hoang, Chernomor, ... Vinh, 2018; Kalyaanamoorthy, Minh, ... Jermini, 2017). The ModelFinder program in IQTree, uses an initial parsimony tree, to determine the best evolutionary model to describe the alignment. For the Cut-PETase alignment the model LG+F+I+G4 was considered the best-fitting by both the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) (final log likelihood of tree is -13117.730) (Le and Gascuel, 2008; Susko and Roger, 2020). Similarly for the Cut-Therm alignment, best fit for Corrected AIC and BIC was model LG+F+I+G4 (final log-likelihood of tree is -16690.591). Resampling of both trees with a bootstrap of 1000 was done simultaneously by the program. Ancestral sequence reconstruction using the same evolutionary model as above (LG+F+I+G4) was completed by the maximum likelihood method in IQTree (Nguyen, Schmidt, ... Minh, 2015). All sequences, alignments, and trees were analyzed using a combination of python, Jalview, and FigTree (Rambaut, 2018; Waterhouse, Procter, ... Barton, 2009). Further ancestral sequences were assessed by AGGRESCAN, and ProtParam (Instability Index, Aliphatic Index) (Conchillo-Solé, de Groot, ... Ventura, 2007; Gasteiger, Hoogland, ... Bairoch, 2005).

2.1.2 Homology Modelling and Analysis

Several homology models for ancestors were generated using template based methods: SWISS-Prot, Alpha Fold, ITasser, Phyre, Robetta CM, and Robetta TTA (Baek, DiMaio, ... Baker, 2021; Bonneau, Tsai, ... Baker, 2001; Jumper, Evans, ... Hassabis, 2021; Kelley, Mezulis, ... Sternberg, 2015; Waterhouse, Bertoni, ... Schwede, 2018). These were compared to crystal structures from Tf. Cut II (PDB 4CG1) and Is. PETase (PDB 6EQE). Models were further evaluated using FoldX Stability and Rosetta (Score_jd2) energy scores, and MolProbity Ramachandran scores (Alford, Leaver-Fay, ... Gray, 2017; Park, Bradley, ... DiMaio, 2016;

Schymkowitz, Borg, ... Serrano, 2005; Williams, Headd, ... Richardson, 2018). Ultimately, the Alpha Fold Models were used for molecular dynamics, and for further analysis by calculation of absolute contact order (ACO), long range order (LRO), and Camsol Structurally Corrected Solubility (Broom, 2016, p. 201; Gromiha and Selvaraj, 2001; Ivankov, Garbuzynskiy, ... Finkelstein, 2003; Sormanni, Aprile, and Vendruscolo, 2015).

2.1.3 AMBER All-Atom Molecular Dynamics

Protein models were obtained from available crystal structures (Tf. Cut II: PDB 4CG1, Is. PETase: PDB 6EQE), or alpha-fold (ANC Cut-Therm, ANC Cut-PETase). For crystal structures, amino acids related to the expression vector for Tf. Cut II and Is. PETase were removed (shown below).

Tf. Cut II: (N-terminal - YPNSSSVDKLAAALEHHHHHH)

Is. PETase (C-terminal - MNFPRASRLMQAAVLGGLMAVSAAATAQT,
N-terminal LEHHHHHH)

Modeller loop modelling was further used to re-insert missing residue, D246, into the Tf. Cut II structure (Šali and Blundell, 1993). After obtaining structures, proteins were pre-prepared using the Schrodinger Maestro preparation wizard, which included prediction of protonation states for ionizable amino acids. Then, Tleap was used to centre structures inside a 12x12x12 Å OPC water box, to complex disulfide bonds, add hydrogen atoms, and neutralize system by addition of 0.1 M Na⁺/Cl⁻. All generated coordinate and topology files were created using the following force fields: AMBER-ff19SB (protein), gaff2 (organic molecules), OPC (water), and ions11m_126_iod_opc (ions) (Tian, Kasavajhala, ... Simmerling, 2020).

Molecular dynamics simulations were performed with the pmemd.cuda engine in the AMBER 20 suite (Case, Belfon, ... Cheatham, 2020). Simulations consisted of four stages: (1) energy minimization of solvent and protein, (2) heating, (3) NPT equilibration, and (4) 100 ns production simulations. Energy minimization was carried out for 5000 cycles with a reduction in protein harmonic constraints every 1000 cycles (from 100 to 5 kcal/mol). Finally, all restraints were removed, and another 1000 cycles of energy minimization was performed. Importantly, each stage of gradient minimization was composed of 100 steps of steepest descent minimization followed by 900 steps of conjugate gradient minimization. Following minimization, the system is gradually heated to the desired temperature over a duration of 100 ps, with control of temperature with a Langevin thermostat. Then, NPT ensembles (controlled using the Langevin thermostat, and Berendsen barostat) with SHAKE and periodic boundary conditions, are used for all equilibration and production simulations. Each equilibration stage spanned a time of 80 ps, for a total equilibration time of 640 ps (8 stages). Stages 1-4 were completed with constraint to the protein (non-hydrogen atoms) (reduced sequentially from 10, 5, 1, 0.1 kcal/mol). Then, in stages 5-6 constraint was placed solely on backbone atoms (CA, C, N) at 0.1 and 0.01 kcal/mol respectively. In order to promote catalytic triad equilibration, stage 7 consisted of a weak constraint to catalytic interactions (0.01 kcal/mol). Finally, the final equilibration stage is run with no constraints. Equilibrated systems are then subjected to productions of 100 ns using an integration time of 2 fs. Simulations were completed in triplicate. All simulations are run at 30 °C unless otherwise stated (this temperature was chosen to match experimental temperature conditions—higher temperatures were tested but data is not reported on for this thesis).

Analysis on simulations is completed by a mixture of CPPTRAJ (RMSF, RMSD, distances, etc.), Visual Molecular Dynamics (VMD, for visualization), and python (for graphing

and data management) (Case, H. Metin Aktulga, ... Kollman, 2022, p. 2; Humphrey, Dalke, and Schulten, 1996). Further, CASTP was used for analysis of simulation frames for calculation of pocket volumes (Tian, Chen, ... Liang, 2018).

2.2 Experimental

2.2.1 Purification of wildtype Tf. Cut II

The codon-optimized wild-type Tf. Cut II, cloned into the pET-26b(+) at restriction enzyme sites NdeI and XhoI, was recombinantly expressed in *E. coli* BL21 DE3 cells (plasmid obtained from GENSCRIPT). Cells were inoculated into 1 L 2TY, with 50 µg/mL Kanamycin, at 37 °C with shaking (200 RPM), until an OD₆₀₀ of ~0.8 was reached. Protein production was induced with addition of 0.1 mM IPTG. Following induction, cells were grown at 25 °C for 8 hours, prior to harvesting by centrifugation at 10,000 g for 20 minutes (4 °C), and resuspension in 100 mM Tris, 10 mM EDTA, 1 M NaCl, pH 8.

Translated sequence:

MANPYERGPNTDALLEARSGPFSVSEENVSRLSASGFGGGTIYYPRENNYGAVAISPG
YTGTEASIAWLGERIASHGFVVITIDTITLQDPSRAEQLNAALNHMINRASSTVRSRID
SSRLAVMGHSMGGGSLRLASQRPDLKAAIPLTPWHLNKNWSSVTVPTLIIGADLDTIA
PVATHAKPFYNSLPSSISKAYLELDGATHFAPNIPNKIIGKYSVAWLKRFVDNDTRYTQF
LCPGPRDGLFGEVEEYRSTCPFLEHHHHHHH

Prior to lysis, cells are incubated in lysozyme and DNase I for roughly 20 minutes on ice. Then lysis was completed using the Emulsiflex®-C5 High Pressure Homogenizer (AVESTIN) at pressures >17,000 psi for 5-10 minutes until a less viscous, less opaque lysate is obtained. Following lysis, supernatant was separated by centrifugation at 16,000 g for 21 minutes (4 °C).

Tf. Cut II was separated from supernatant by nickel immobilized metal affinity chromatography using Profinity IMAC resin (Biorad Laboratories Inc) and a BioRad low pressure chromatography system (BioRAD BioLogic LP, BioRad Laboratories Inc). Following supernatant loading, protein was washed with cooled wash buffer (300 mM NaCl, 20 mM sodium phosphate, 20 mM imidazole, pH 7) at 2 mL/min until the baseline returned to normal (roughly 2-3 hours).

Then, protein was eluted with 300 mM NaCl, 50 mM sodium phosphate, and 150 mM imidazole, pH 7 at 1 mL/min. Eluted protein was dialyzed thrice with either 50 mM Tris, pH 8, 100 mM NaCl or 50 mM ammonium carbonate (~pH 8.9) depending on need for lyophilization. Finally, Tf. Cut II was concentrated to 10-12 mg/mL using an Amicon® stirred cell (EMD Millipore Corporation) with a molecular weight cut-off filter of 10 kDa. After this, proteins were stored in solution or as lyophilized powder at -80 °C.

2.2.2 Purification of Ancestor Proteins

Similar to the wild-type Tf. Cut II, codon optimized ancestors were inserted into pET-26b(+) (ANC Cut-PETase, NdeI-XhoI) or pET-29b(+) (ANC Cut-Therm, NdeI-XhoI), by Genscript and Twist Biosciences respectively. Obtained expression vectors, were cloned into *E. coli* BL21 DE3 cells. These cells were inoculated into 1 L 2TY with 50 µg/mL Kanamycin at 37 °C with shaking (200 RPM), until an OD of ~0.8 was reached, after which protein expression was induced with 0.1 mM IPTG. Expression was continued for 16 hours at 25 °C. Following expression, cells were harvested by centrifugation at 4000 g for 20 minutes (4 °C), resuspended in 100 mM Tris, 10 mM EDTA, 1 M NaCl, pH 8, and stored at -80 °C.

Translated sequence ANC Cut-Therm:

MNPYERGPNPTESSIEALRGPFSVDEERSRLAASGFGGGTIYYPTDNNEGTFGAVAIISP
GYTGTQSSISWLGERLASHGFVVMTIDTNTTLDQPDSRASQLDAALDYMVVELSSDSSSS
VRNRIDSSRLAVMGHSMGGGGTLRLAERRPDLKAAIPLTPWHLDKTWSSVRVPTLIIGA
ENDTIASVRSHSEPFYNSLPGSLDKAYLELDGASHFAPNLSNTSSGTIAKYSISWLKRFVD
DDTRYTQFLCPGPSTGLFSDVEEYRSTCPLEHHHHHH

Translated sequence ANC Cut-PETase:

MNPYERGPDPTESSIEALRGPFSVDEESVSSLAVSGFGGGTIYYPTDTNEGTFGAVAIAPG
YTASQSSMSWLGERLASQGFVFTIDTNRRLDQPDSRGDQLSAALDYLVEDNGRSSSAV
RNRIDPDLAVMGHSMGGGGTLAAAEDRPSLKAAIPLTPWHLDKTWSEVRVPTLIIGAE

NDTIASVRSHSEPFYNSLPGSLDKAYLELDGASHFAPNLSSSYNTTIAKYSISWLKRFVDD
DTRYSQFLCPGPSTGLGSDVEEYRSTCPLHHHHHH

Cell lysis was completed by emulsification as described for Tf. Cut II or by freeze-thaw. In the freeze-thaw method, lysozyme (> 1 mg/mL) and DNase are added to cells for 30 minutes with shaking. Then, 10 rounds of freeze-thaws with inter-mixed vortexing are completed to lyse cells. Pellet is separated from supernatant by centrifugation at 10,000 g for 30 minutes (4 °C). The separated pellet is resuspended overnight with spinning in urea denaturation buffer (100 mM sodium phosphate, 10 mM Tris, 7 M Urea, pH 8.1).

Solubilized pellets are centrifuged at 16,000 g for 30 minutes (4 °C) to remove undesired cellular components. Then supernatant is filtered using vacuum filtration with a slow-flow filter. This supernatant is applied to Profinity IMAC resin as described for Tf. Cut II. One difference is that wash and elution buffers both contain 4 M urea (wash buffer is 300 mM NaCl, 4 M Urea, 100 mM Tris, 20 mM imidazole, pH 7.5; elution buffer is 300 mM NaCl, 4 M Urea, 100 mM Tris, 100 mM Imidazole, pH 7.5).

Protein refolding was first tested by fluorimetry, and aggregation based light scattering on a Horiba fluorimeter (model FL3-11). First, 50 uL of unfolded protein was diluted into 1000 uL pre-equilibrated refolding buffer (final concentration around 1.06 μM), and light scattering was measured for 5 minutes (1 cm pathlength, 450 em, 450 ex, 27 °C). Scattering measurements—particularly the gain of scattering from pre-protein addition may reflect aggregation of ancestors in solution. Following scattering measurements, Trp fluorescence spectra were collected by excitation at 280 nm, and collection at 305-400 nm. Blue shifting of spectra from the unfolded state was considered to be a reflection of Trp burial, and indicative of folding.

Ultimately, dialysis and refolding of ancestors was completed thrice in 50 mM ammonium carbonate (~ pH 8.9) following incubation of unfolded protein with 10 mM dithiothreitol (DTT) and 0.45 mM EDTA. Finally, ancestors were concentrated to 4-5 mg/mL using a combination of AMICON® stirred cell, and AMICON® ultra-15 centrifugal filter units (centrifuged in spinning rotor at 4,000 g, 20 °C).

Size exclusion chromatography (SEC) was routinely performed to reduce the presence of higher-order disulfide bonded protein species. This was completed in large scale using a High Load 60/600 Superdex 200 column (Cytiva) in 200 mM ammonium carbonate (this concentration helps reduce ancestor aggregation, and have the minimum requirement of 0.15 M salt presence for optimal column performance). The ammonium carbonate solution was filtered via vacuum filtration, and degassed with nitrogen gas for 20 minutes prior to use. As the sample typically elutes as one broad peak (which may be due to aggregation), the peak is collected in small volume fractions to improve chances of resolving different protein species.

2.2.3 Chemical Denaturation Curves

Protein stocks were diluted into solutions of varying urea (ancestors) or guanidine hydrochloride (Tf. Cut II) concentrations in 50 mM sodium phosphate, 100 mM NaCl, pH 7.5. All volume additions were measured by mass on an analytical balance to ensure variation of sample protein concentrations was kept to a minimum (<2.5 % variance). The diluted protein solutions were incubated at 27 °C for 30 minutes. Then, solutions were transferred and equilibrated in a 1 cm cuvette, within the fluorimeter (Horiba model FL3-11) at 27 °C. Following a 2-minute equilibration period, tryptophan spectra was collected by excitation at 280 nm and collection at

305-400 nm. Following collection, samples were transferred to a secondary cuvette and incubated with 4 mM DTT before remeasurement.

Intensity of unfolding was measured for Tf. Cut II at 327 nm, and plotted against increasing guanidine hydrochloride concentration to create the denaturation curve. As intensity in ancestors is unable to be used for this purpose, maximum wavelength was plotted against increasing urea concentration instead. In this case, smoothing of each spectrum was required, and was completed by a LOESS smoothing function in GRAPHER (GOLDEN Software). Finally, resulting denaturation curves were fitted to a linear-extrapolation model (Equation 2) using OriginLab.

$$Y = (Y_N + S_N [Dn]) - \frac{((Y_N + S_N [Dn]) - (Y_U + S_U [Dn])) \left(e^{\frac{\Delta G_{U-F} - m_{eq} [Dn]}{RT}} \right)}{1 + e^{\frac{\Delta G_{U-F} - m_{eq} [Dn]}{RT}}} \quad (2)$$

where Y is the intensity or maximum wavelength of the native (Y_N) or unfolded (Y_U) state, S is the slope of native (S_N) or unfolded state (S_U) when denaturant (Dn) is present, ΔG_{U-F} is the free energy of unfolding in water, m_{eq} is related to denaturant dependent unfolding, R is the gas constant (1.987 kcal mol⁻¹ K⁻¹), and T is the temperature in Kelvin (set to 300.15 K, or 27 °C). All protein denaturation curves seem to fit well to the above model.

2.2.4 Circular Dichroism

Circular dichroism (CD) was performed using a JASCO-715 with an 0.01 cm path length cuvette. Protein samples were added to the cuvette at a final concentration of 1 mg/mL with or without 5 mM reductant tris(2-carboxyethyl)phosphine (TCEP). Spectra was collected from 260 to 185 nm, at a scanning speed of 50 nm/min (bandwidth = 2.0 nm, accumulations = 3, data pitch = 0.2 nm, response = 2 sec, 27 °C). All collections were completed in 10 mM sodium phosphate

buffer, pH 7.5. Spectral units (ΔA in degrees) were converted to mean residue (MR) ellipticity $[\theta]$ ($\text{deg cm}^2 \text{ dmol}^{-1}$) using Equation 3.

$$[\theta]MR = \frac{100 * \Delta A * 32.982}{CMR * l * 0.1} \quad (3)$$

$$CMR = Conc (M) * N \quad (4)$$

where CMR is the mean residue concentration (calculated by multiplying protein concentration in molar with number of amino acids (N)), l is the cuvette pathlength, and ΔA is the optical signal measured by CD in degrees.

Secondary structure prediction was completed using BeStSel which is an improved method for analyzing proteins with high beta content (Micsonai, Moussong, ... Kardos, 2022). On top of normal secondary structure predictions, it further provides a fold class and related PDB structures to the provided spectra.

2.2.5 ATR-FTIR

Attenuated total reflectance Fourier transform infrared spectroscopy (ATR-FTIR) was completed as described (Naser, Tarasca, ... Meiering, 2022). Briefly, a Tensor 37 FTIR spectrometer (Bruker Optics), equipped with a mercury-cadmium-telluride detector (spectrometer under continuous N_2 purge, and N_2 (l) cooling of detector) was used to acquire protein spectra. For each sample, four spectra are consecutively collected (256 scans each, range of $4000\text{-}1200 \text{ cm}^{-1}$). Prior to protein collection, spectra for background (air) and buffer (flow-through from purification) are acquired. Further, following protein collection, and flushing of crystal with buffer, another buffer spectra was re-collected (post-buffer). All measurements were made at $25 \text{ }^\circ\text{C}$.

Data processing of FTIR spectra was completed using the OPUS 6.5 software (Bruker). Data processing included atmospheric correction and post-buffer subtraction. Here, subtraction of post-buffer from protein sample measurement may help account for signal from protein aggregates cemented on the crystal.

2.2.6 *pNPA and pNPP activity*

Esterase activity for Tf. Cut II and ANC using *p*-nitrophenyl acetate (PNPA) and *p*-nitrophenyl palmitate (pNPP) was measured. Both products form *p*-nitrophenol upon hydrolysis, a yellow product that can be seen visually and by absorbance measurements at 405 nm. Thus, all pNPA measurements were completed on a microplate reader (SpectraMax M2, SoftMax Pro 6), while pNPP measurements were completed visually (due to insoluble precipitate formation obscuring absorbance readings).

Foremost, changes to Tf. Cut II rate ($6.9\text{E-}3 \mu\text{M}$ in 50 mM sodium phosphate, 100 mM NaCl, pH 7.5) against $1.25 \mu\text{M}$ pNPA dissolved in 2.5-10% ethanol, DMSO, and acetonitrile were measured. Based on these results, and substrate solubility, pNPA measurements were done at a final 2.5% ethanol percentage, and all pNPA measurements were done at a final 5% DMSO percentage.

For ancestors, multiple protein concentrations (ranging from dilute to 4-5 mg/mL) were tested for increases to pNPA hydrolysis (various concentrations tested) above background rate. However, no rate above background rate could be observed. Further, incubation of concentrated ancestors (4-5 mg/mL) with pNPP ($0.1 \mu\text{M}$) resulted in no color change even after several hours of observation. Wild-type measurements against pNPP was completed in a similar manner, to act

as a positive control. The concentration of pNPP was kept low (0.5-1 mM) to avoid overproduction of insoluble precipitate.

Michaelis-Menton kinetics for Tf. Cut II was measured such that the final assay mixture contained 195 uL enzyme buffer solution (6.9E-3 μ M final concentration in 50 mM sodium phosphate, 100 mM NaCl, pH 7.5), and 5 uL substrate (0-2.25 mM pNPA in Ethanol, final ethanol % in assay of 2.5). Rate was measured in triplicate by increases to absorbance at 410 nm at all substrate concentrations. Importantly, non-enzymatic hydrolysis of pNPA with buffer was also measured and subtracted to obtain a corrected rate. Finally, Michaelis Menten fit was obtained using SOLVER (Microsoft Excel) using Equation 5.

$$V_0 = \frac{V_{\max} [S]}{K_m + [S]} \quad (5)$$

where V_0 is the initial rate of the reaction, V_{\max} is the maximum initial rate, $[S]$ is substrate concentration, and K_M is the Michalis-Menten constant.

Residual activity was determined by initial rate retention against pNPA following incubation of dilute enzyme (5.56 μ g/mL) at increasing temperatures (32-80 °C). Importantly, following a 30-minute incubation, samples were placed on ice for 5 minutes, before measurement of rate (completed in duplicate). Initial rates were measured for a final pNPA concentration of 2.2 mM, and a final protein concentration of 3.9 nM.

2.2.7 Mass Spectrometry

Mass spectrophotometry was completed by the mass spectrometry facility at the University of Waterloo. In preparation for mass spectrometry ANC Cut-PETase was buffer exchanged into

water four times using an AMICON® ultra-15 centrifugal filter unit and concentrated to 1.13 mg/mL (centrifuged in spinning rotor at 4000 xg, 20 °C). Then, Valerie Goodfellow aided with running the ancestor on linear mode MS using a Bruker Autoflex Speed matrix assisted laser desorption ionization (MALDI) spectrometer and a 3-hydroxypicolinic acid matrix. Standards with molecular weights 14.17, 19.97, 23.98, and 29.02 kDa were also included to create a standard curve.

2.2.8 Iodoacetamide Gel

In order to verify or examine the state of the disulfide following chemical denaturation experiments, iodoacetamide labelling of reduced Cys residues was completed. Frozen experiment protein samples were thawed and precipitated by a 20-minute incubation with equal volumes of 20% trichloroacetic acid. Samples were pelleted by centrifugation at 14,000 g for 15 minutes at 4 °C, following which the supernatant was separated. The pellet was washed by resuspension in 100 µL of acetone, followed by recovery with a second centrifugation step (14, 000 g, 15 minutes, 4 °C). The acetone supernatant was removed, and the pellet was dried under vacuum. The dried pellet was redissolved in 50 µL of 50 mM Hepes (pH 7.8), 2.5 % sodium-dodecyl sulfate, 1 mM bathocuproine disulfonate, and 100 mM iodoacetamide (final pH of 7.2). The sample was incubated for 1 hour at 37 °C in a desiccator. Finally, samples were run in non-reducing conditions on SDS-page.

3 Results

3.1 A rich sequence space leads to important, early choices

The alpha-beta hydrolase fold could be considered to be a super-scaffold, as many sub-families with varying substrate specificity have emerged within it. Indeed, following initial BLAST results, over 7000 sequence hits were obtained with diverse annotations (cutinases, hydrolases, lipases, etc.). In order to perform ASR in this vast class of enzymes, it became pertinent to fish out enzymes that were more likely to produce PET hydrolase like properties. Thus, two known PET-degrading enzymes, Tf. Cut II and Is. PETase were used as bait. During collection by BLAST, sequences were dismissed if they presented as too divergent from the target enzymes (<30% sequence identity), having subpar coverage (<75% of Tf. Cut II length), or were identical to another sequence by more than 95% (in which case, only one was kept). At this stage, 900 sequences remained, implying that this approach was sufficient to filter away >6000 sequences. Hereafter, filtering of the remaining sequences became complicated and there were several options that were considered rationally adequate: (1) the range of percent identity allowed could be narrowed (say between 50-75%), (2) the closest Is. PETase or Tf. Cut II relatives could be selectively used, or (3) sub-classes for enzymes could be chosen based on annotation data. Ultimately, there were concerns for both options (1) and (2), given that known PET hydrolases themselves only share 50% sequence identity, yet display PET degrading activity. It was likely that overall sequence identity would not be a sufficient metric to capture valuable activity signal. Secondly, there was concerns of only using sequences with high relation to Is. PETase, as it is considerably less stable than Tf. Cut II and LCC. The third option, while reliant on correct annotations of enzymes, seemed optimal to retain known thermostable and active extant proteins.

Thus, multiple sequence alignments were created using sequences amongst the 900 that were labelled as cutinases (24), thermostable (25), and lipases (25) were chosen based on similarity to Tf. Cut II) (MSA Cut-PETase, APPENDIX B). Further, the second alignment (MSA Cut-Therm, APPENDIX C) contained a group of sequences (25) with the closest sequence identities to Is. PETase.

Examination of the alignments—specifically with comparison to the Is. PETase structure—reveals that the majority of residue variability is seen on surface exposed residues. Residues related to inner beta sheet formation and connections thereof have conserved properties (>98% similarity). Residues related to esterase activity, such as the G-X-cS-X-G motif, and the catalytic triad is likewise conserved. Hence, there is an expectation that the majority of changes in the predicted ancestors will be on loops, surface exposed residues, and at the N-terminal (which is surface exposed and unstructured). Importantly, the C-terminal displays high variation with regards to residue identity and length. It presents an interesting case, as the end section of the C-terminal forms a final, but less conserved beta-strand, which connects to an upstream alpha helix by a relatively long loop. Due to the high flexibility of this loop (as seen in the b-factors of PDB 4CG1, 3VIS, 4WFI), it is suspected that the C-terminal disulfide, which directly links the C-terminal beta-sheet to the alpha helix plays a critical role with the stabilization of this area. As mentioned previously, loss of this disulfide in LCC, generates less stable variants (Δ apparent T_m of -16 °C), though it does not greatly affect activity (Sulaiman, You, ... Kanaya, 2014). In order to navigate the alignment of this region, the length of the C-terminal loop was set to match enzymes from the Thermobifida family—in hopes that this length, which is seen in thermostable enzymes, would be adequate for stable ancestor generation. Additionally, the extra loop regions of Is. PETase were retained, based on the prose that the flexibility obtained by these loops is highly

relevant to PET binding (Joo, Cho, ... Kim, 2018). One of these is situated on the cHis loop as previously described (Loop Insert #2), while the other extends a conversely positioned loop between a beta sheet and alpha helix (Loop Insert #1). In the alignments, specifically for the PET-Cut alignment, these sites are not well occupied, and present a possible risk with their inclusion. Further, the effect of removal of either loop insert has not yet been experimentally characterized in *Is. PETase*.

3.2 Locating desirable ancestral nodes for reconstruction

All known genomes contain some member of the alpha/beta hydrolase family; the adjustability of this fold allows it to be nature's answer to a number of catalytic problems (Marchot and Chatonnet, 2012). Thus, its evolution is eons long, not simple to determine, and is convoluted by multiple gene copies evolving for separate functions. The phylogenetic trees created henceforth, are likely not representative of the true evolution of cutinases or lipases, but rather hope to capture characteristics relevant to PET acceptance. An extremely important observation that is immediately noticed in both alignment trees (Figure 3.2, Figure 3.3) is that separation of clades is not related to annotated function (i.e. lipases are found through-out the tree, and are not singly localized). Instead, grouping tends to be more related to originating species. This may in some cases be related to naïve labelling of function (based on sequence identity, rather than experimental testing), but it does represent the difficulty in pinpointing the evolutionary path for these multi-substrate accepting enzyme families.

Surface charge has been considered as a mechanism for which PET-film/enzyme interaction can be modulated (Furukawa, Kawakami, ... Miyamoto, 2019, 2018; Sagong, Son, ... Kim, 2021). The predicted pI for sequences in the tree suggest that the majority of the tree contains acidic proteins (pIs of between 4-6) (APPENDIX E). This is contrary to the high pIs of Is. PETase and LCC (~9.5). As the mechanism for enzyme approach to PET-film binding is still unclear, the potential benefits of a basic pI are not confirmed.


It could be argued that PET hydrolyzing activity in the alpha-beta hydrolase class is not due to deterministic evolution—i.e., these enzymes have not had sufficient time, nor requirement to evolve specificity towards PET degradation, and thus, any ability to degrade PET is a result of

substrate promiscuity. At this time, Is. PETase has the fastest natural rate for PET degradation at low temperature, yet embodies many unique features that could contribute to its reduced stability (Lu, Diaz, ... Alper, 2022; Yoshida, Hiraga, ... Oda, 2016). On the other hand, LCC (48.5% identity to PETase), overcomes the activity seen in Is. PETase at higher temperatures (Tournier, Topham, ... Marty, 2020, p. 202; Yoshida, Hiraga, ... Oda, 2016). Thus, an ancestor node that includes both Is. PETase and LCC may be particularly interesting. In the alignment discluding the close PETase homologs (Cut-PETase) (Figure 3.2), a clade containing both LCC and Is. PETase was predicted, however only contained a total of 3 sequences. Upstream of this clad, was a family of thermostable cutinases that have all reported some PET degradation activity. Despite a lower support for this connection (bootstrap = 660/1000), there was an interest in the minutaie this node could be capturing. The resurrection of this ancestor, called ANC Cut-PETase was thus undertaken. Additionally, a second ancestor for the thermostable cutinase clade of the Cut-Therm phylogenetic tree was resurrected. This ancestor was expected to exhibit high thermostability, and promiscuity akin to its extant enzymes. Ultimately, two ancestors were chosen for reconstruction to obtain initial feedback on the design of the MSA and tree topology prior to continuing with auxillary nodes (Figure 3.1).

```

>Cut-PETase -NPYERGPDPTESSIEALRGPFSVDEESVSSLAVSGFGGGTIYYPTDNEGTFGAVAIAPGYTASQSSMSWLG
>Cut-Therm -NPYERGNPTESSIEALRGPFSVDEERVSRLAASGFGGGTIYYPTDNEGTFGAVAIAPGYTGTQSSISWLG 72
>Tf. Cut II ANPYERGNPTDALLEARSGPFSVSEENVSRLSASGFGGGTIYYPREN--NTYGAVAIAPGYTGTTEASIAWLG


```



```

>Cut-PETase ERLASQGFVVFTIDTNTRLDQPDSRGDQLSAAALDYLVEDNGRSSSAVRNRIDPDR LAVMGHSMGGGGTLAAA
>Cut-Therm ERLASHGFVVMTIDTNTTLDQPDSRASQLDAALDYMVELSSDSSSSVRNRIDSSRLAVMGHSMGGGGTLRLA 144
>Tf. Cut II ERIASHGFVVITIDTITTTLDQPDSRAEQLNAAALNHMIN---RASSTVRSRIDSSRLAVMGHSMGGGGSLRLA


```



```

>Cut-PETase EDRPSLKAAIPLTPWHLDKTWSEVRVPTLIIGAENDTIASVRSHSEPFYNSLPGSLDKAYLELDGASHFAPN
>Cut-Therm ERRPDLKAAIPLTPWHLDKTWSSVRVPTLIIGAENDTIASVRSHSEPFYNSLPGSLDKAYLELDGASHFAPN 216
>Tf. Cut II SQRPDLKAAIPLTPWHLNKNWSSVTVPTLIIGADLDTIAPVATHAKPFYNSLPSISKAYLELDGATHFAPN

```



```

>Cut-PETase LSSSYNTTIAKYSISWLKRFVDDDDTRYSQFLCPGPSTGLGSDVEEYRSTCP-
>Cut-Therm LSNTSSGTIAKYSISWLKRFVDDDDTRYTQFLCPGPSTGLFSDVEEYRSTCP- 267
>Tf. Cut II IPNK---IIGKYSVAWLKRFVDNDTRYTQFLCPGPRDGLFGEVEEYRSTCPF

```




Figure 3.1 Sequences of ancestors Cut-PETase and Cut-Therm. Sequences predicted by ancestral sequence reconstruction for nodes connecting thermostable cutinases and Is. PETase group (Cut-PETase), and for a thermostable cutinase clade (Cut-Therm) are shown. The sequence for Tf. Cut II is directly below for reference. Predicted secondary structure of sequences are shown, such that arrows and squares represent alpha helix and beta sheets respectively. Loops are shown as lines connecting secondary structures. Sites of loop inserts are bolded and underlined (two loop inserts). Finally, catalytic residues are coloured green and identified by an upwards pointing arrow. The conserved disulfide bond Cys residues are located at the C-terminal (in red). The final length of the ancestor proteins is 267 residues.

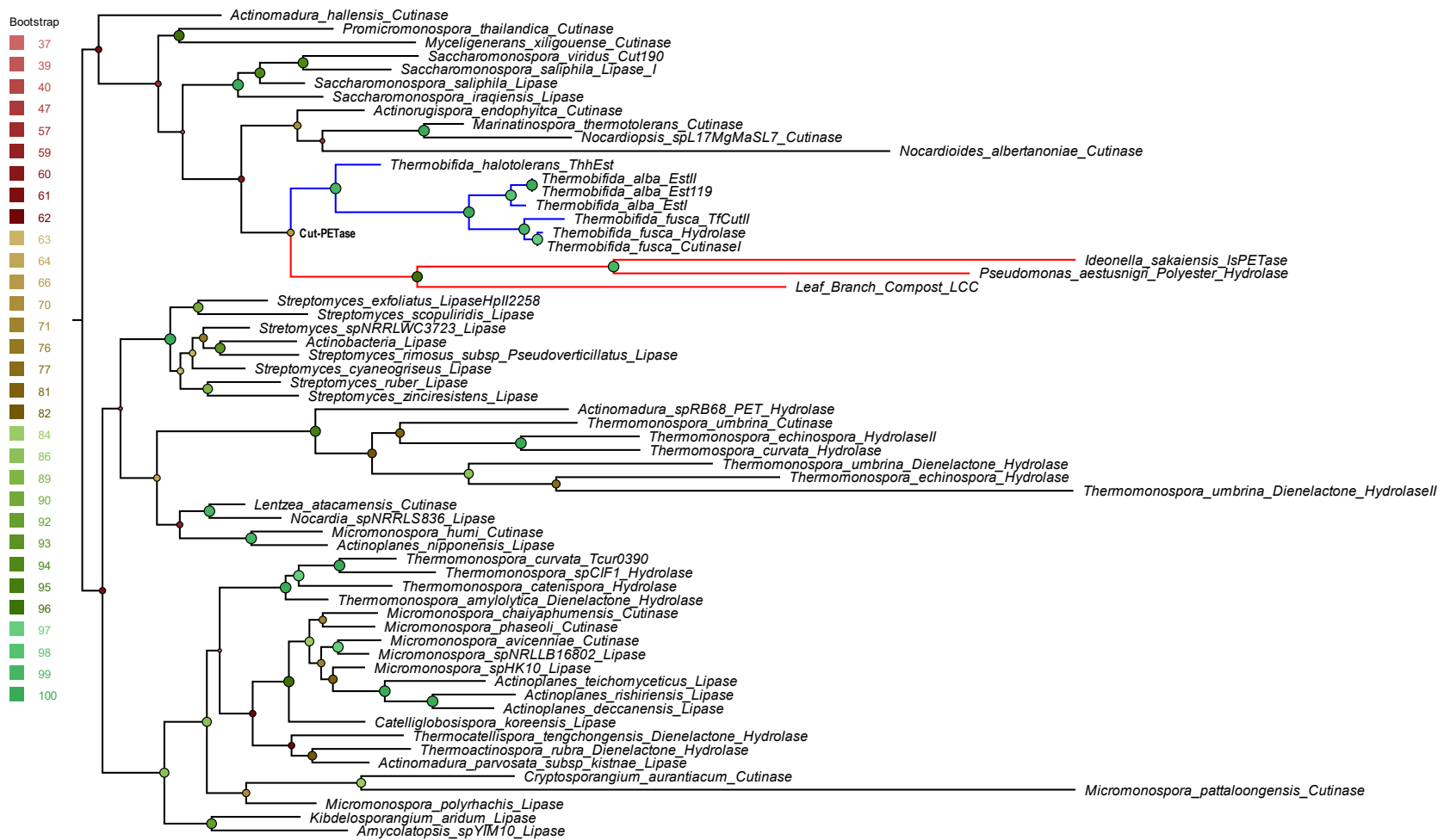


Figure 3.2 Cut-PETase Phylogenetic Tree The above tree contains sequences from the cutinase, thermo, and lipases groups, as well as known PET hydrolases. Is. PETase and LCC are located in the red clade, while the blue clade hosts thermostable cutinases (including Tf. Cut II). The reconstructed ancestor, Cut-PETase belongs to the node shared between them (labelled Cut-PETase, bootstrap value = 66%). Bootstrap percents (sampling = 1000 times) are shown as circles at nodes, where higher bootstrap values are represented as larger circles. Further, circles are coloured based on bootstrap percent as dictated by the legend (left). This tree was created using IQTree, with evolutionary model LG+F+I+G4. Figure was made using FigTree, with midpoint rooting of tree.



Figure 3.3 Cut-Therm Phylogenetic Tree The above tree contains sequences from the cutinase, thermo, PETase homologs, and lipases groups, as well as known PET hydrolases. Is. PETase and LCC are located in the red clade, while the blue clade hosts thermostable cutinases (including Tf. Cut II). Sequences that are unique to the Cut-Therm tree, are coloured red, while the remaining are shared between both Cut-Therm and Cut-PETase trees. A more conserved site was chosen for initial reconstruction (labelled Cut-Therm, bootstrap value = 100%). Bootstrap percents (sampling = 1000 times) are shown as circles at nodes, where higher bootstrap values are represented as larger circles. Further, circles are coloured based on bootstrap percent as dictated by the legend (left). This tree was created using IQTree, with evolutionary model LG+F+I+G4. Figure was made using FigTree, with midpoint rooting of tree.

3.3 Ancestor expression and purification

Expression of the wild-type Tf. Cut II and ancestors was completed recombinantly in *E. coli* BL21 (DE3) using the T7 expression system. Notably, all expression attempts with varying temperature (16, 18, 25, 37 °C), IPTG concentration (0.05, 0.1, 0.5 mM), media (2TY, minimal, LB) or cell type (BL21 DE3, BL21 DE3 PlyS) on both ancestor proteins did not yield soluble protein. This is in contrast to Tf. Cut II, which was highly soluble (Figure 3.4). It is not uncommon for some proteins to be expressed as inclusion bodies, and even Is. PETase is known for limited soluble production (Jo, 2021; Ko, Kang, ... Sung, 2021; Shi, Liu, ... Zhu, 2021). Further, purification from the insoluble fraction can be beneficial for obtaining highly pure protein, as it is typically less populated by native, host proteins. However, in the consideration of the desired application (i.e., bio-based recycling), good solubility is of paramount necessity, and thus, in their current state these ancestors are not candidates for recycling. Nevertheless, in order to characterize their stability and activity both Tf. Cut II and ancestors were purified from the soluble and insoluble fraction respectively using a combination of nickel affinity chromatography and size exclusion chromatography (Figure 3.4).

The predicted molecular weights of the proteins are very similar: Tf. Cut II (29.29 kDa), Cut-PETase (29.89 kDa), and Cut-Therm (30.07 kDa). Yet, on SDS-page the calculated molecular weight of the ancestors is around 35 kDa (Figure 3.4). Accordingly, mass spectrometry on the Cut-PETase ancestor, which confirmed expression of Cut-PETase with a molecular weight of 29.91 kDa, was necessary to corroborate the expression of the expected protein sequence (APPENDIX F). This variation between predicted molecular weight and SDS-page molecular weight has been observed previously for acidic proteins (Guan, Zhu, ... Peng, 2015). This is colloquially explained by acidic domains repelling SDS, the primary driving molecule for protein movement along the

gel (Tiwari, Kaila, and Guptasarma, 2019). Both ancestors are predicted to have a pI of around 5 (compared to Tf. Cut pI of 6.7), which may cause poorer SDS binding and migration towards the anode.

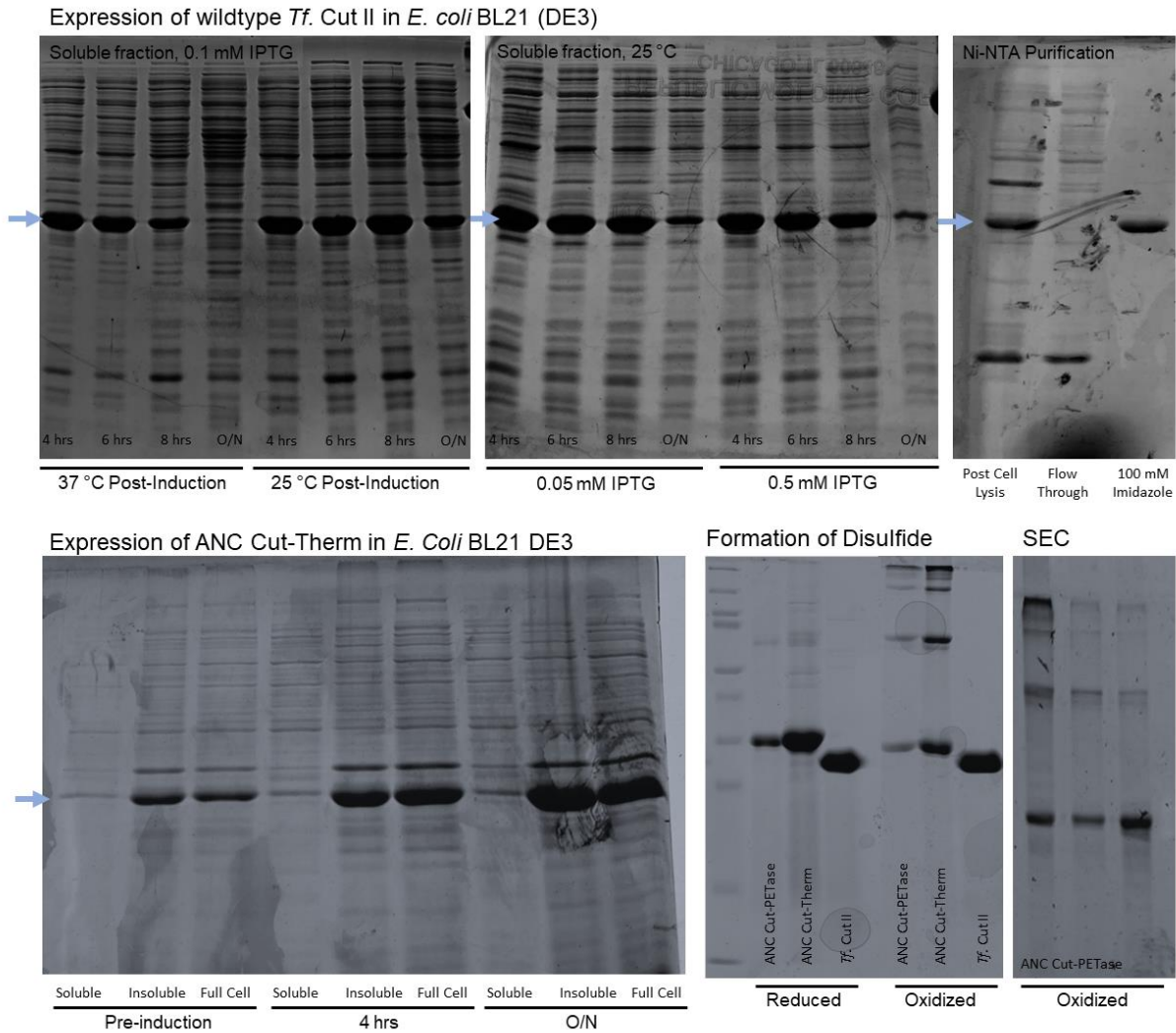


Figure 3.4 Expression of *Tf. Cut II*, ANC Cut-Therm, and ANC Cut-PETase. Expression of *Tf. Cut II* was optimized for temperature and IPTG concentration. Blue arrow corresponds to band of interest. All gels are composed of 15% resolving, and 5% stacking components. **(Top Left)** Gel shows expression levels of *Tf. Cut II* (blue arrow) grown at 37 °C or 25 °C post-induction (0.1 mM IPTG, LB media). Results suggest that expression is higher at 25 °C. **(Top Centre)** Gel shows expression levels of *Tf. Cut II* (blue arrow) grown at 25 °C following induction with 0.05 mM IPTG or 0.5 mM IPTG (LB media). Results suggest better long term expression at 0.05 mM IPTG. Thus, optimal conditions for expression were determined to be at 25 °C for 8 hrs (0.1 mM IPTG) in 2TY. **(Top Right)** Gel of *Tf. Cut II* (blue arrow) purification via nickel affinity chromatography shows pure elution band after addition of 100 mM imidazole. **(Bottom Left)** Expression of ANC Cut-Therm in *E. coli* BL21 (DE3) induced with 0.1 mM

IPTG, and grown at 25 °C. Blue arrow corresponds to ancestor band. Lanes are split by soluble, insoluble, and full cell fractions. This result is representative of numerous attempts on improving ancestor solubility, but with no success. Thus, final expression of the ancestors is done overnight at 25 °C following induction with 0.1 mM IPTG in LB or 2TY to maximize inclusion body yield. **(Bottom Center)** Comparison of reduced and oxidized proteins reveal higher order structures forming in oxidized ancestor samples. This suggests a non-trivial portion of the sample is creating non-natural disulfide linkages. The farthest left lane is a pre-stained markers from ThermoFischer, and aids in molecular weight estimation (Spectra™ Multicolor Broad Range Protein Ladder). This marker contains proteins with molecular weights of: 260, 140, 100, 70, 50, 40, 35, 25, 15, and 10 kDa. Bands for ancestors aligns closely to the 35 kDa marker band. **(Bottom Right)** Size exclusion chromatography (SEC) can be used to remove a majority of higher-order structures, and obtain monomeric ancestors. Each lane represents various fractions from SEC, with the leftmost lane eluting earlier, followed by the centre and right lane. For details see [Methods](#).

Intrinsic protein fluorescence from tryptophan, tyrosine, and less commonly phenylalanine has been widely used to monitor changes in local protein structure (Buchner and Kiefhaber, 2005). Briefly, movements of these residues from a hydrophilic environment to a hydrophobic one, as might be seen in native protein folding, results in a measurable blue shift. Accordingly, several buffers were tested via degree of blue shifting of tryptophan as candidates for ancestor refolding (Table 4). In Tf. Cut II and ancestors this tryptophan signal is a combined reflection of the environments of four tryptophan residues (Figure 3.1), with varying degrees of burial: 0, 3, 35, and 42% CAMSOL SASA in Tf. Cut II, where CAMSOL SASA is the exposure of a residue in the enzyme compared to the same residue in a G-X-G peptide (Sormanni, Aprile, and Vendruscolo, 2015). The average and effectively unaltered max wavelength of Tf. Cut II in all buffers is 329 nm, which is not realized by the ancestral proteins (average around 333 nm). There is, therefore one or multiple tryptophan residues in a slightly more hydrophilic environment than in Tf Cut II. Interestingly, one tryptophan residue—SASA 36% in Tf. Cut II, but predicted to be 28% and 50% in ANC Cut-PETase and ANC Cut-Therm respectively—is partly buried by the cHis loop, where ancestors have an extra three residue extension compared to Tf Cut II (APPENDIX G). Hence, it

could be theorized that some variation (but not all) to the max wavelength of ancestors is related to an alternate conformation of the cHis loop.

Attempts to refold the ancestor proteins in any of the buffers used for fluorescence measurements resulted in an excessive loss of protein due to aggregation. Thus, refolding and storing of ancestors was completed in ammonium carbonate, resulting in considerably less aggregation and similar maximum protein wavelengths of fluorescence (332-334 nm). Ancestors were afterwards diluted in sodium phosphate, pH 7.5, 100 mM NaCl for analysis. This buffer was ultimately chosen due to reduced ANC Cut-Therm aggregation, which is more aggregation prone than ANC Cut-PETase (Table 4).

Table 4 Tryptophan fluorescence and light scattering as a metrics to determine optimal refolding buffer with respect to folding and aggregation Lower maximum wavelengths for ancestors suggest a more folded ancestor structure. These values can be compared to a maximum wavelength of 329 nm for Tf. Cut II. Important to interpretation, both ancestors and Tf. Cut II share identical Trp residues locations. Increases to light scattering overnight is presented as a percent change (Δ LS). Higher percents suggest increased aggregation. Notably, ANC Cut-Therm showed more aggressive aggregation compared to ANC Cut-PETase. All additives were tested in 50 mM sodium phosphate, pH 7.5 at 25 °C. [TMAO – trimethylamine N-oxide, NOG – n-oxyalyglycine].

Buffer (50 mM)	ANC Cut-PETase		ANC Cut-Therm	
	Max Wavelength (nm)	Δ LS (%)	Max Wavelength (nm)	Δ LS (%)
Sodium Phosphate, pH 7.5	334.6	+123	333.3	+35
Potassium Phosphate, pH 7.5	335.1	-6	334.4	+1167
Tris, pH 7.5	334.4	+33	333.5	+1689
HEPES, pH 7.5	332.6	+58	332.6	+2219
ADDITIVES				
0.25 M NaCl	334.2	+323	332.6	+122
0.50 M NaCl	333.9	+351	333.1	+215
0.75 M NaCl	332.9	+224	333.3	+363
1 M NaCl	333.9	+131	332.9	+354
0.5 M Na ₂ SO ₄	333.5	+25	332.4	+467
0.5 M Glycine	334.9	+33	334.6	+674
0.25 M TMAO	334.9	-18	334.2	+434
22.5 mM NOG	334.2	+81	334.0	+1908

A common theme that will arise in the remainder of this thesis, is the difficulty the ancestors experienced in C-terminal disulfide formation (Figure 3.4). By subjecting identical

ancestor samples to SDS-page under oxidizing conditions, it was determined that while some portion of the ancestor samples exists in a monomeric state, a considerable percentage have bands correlating with disulfide driven formation of dimers, trimers, and higher. During refolding, disulfides present an additional complication primarily due to (1) non-native disulfide formation, and (2) early formation leading to non-native folding of protein (Tsumoto, Ejima, ... Arakawa, 2003). In order to support formation of the native structure during refolding in ammonium carbonate (~pH 9), DTT and EDTA was added initially to grant time for native formation, which would then be followed by replacement of the reservoir and subsequent dilution of DTT. As seen this method did not fully prevent the formation of non-native bonding, possibly due to the exposed and flexible nature of the C-terminal. Thus, size exclusion chromatography was used to obtain fractions with a monomer majority (referred to as monomeric ancestors) (Figure 3.4).

3.4 Chemical denaturation suggests low stability of ancestors

Many proteins follow a cooperative, two-state unfolding model, in which proteins transition from a native to unfolded state in an “all or none” fashion. This is usually in correlation with the complexity and degree of networking within a protein core (and thus typically increases with increasing protein size) (Broom, 2010). In order to probe the ancestors for cooperativity, which could suggest they are forming a native like structure, changes to Trp fluorescence in urea was measured. Notably, unfolding equilibria for Tf. Cut II was measured in the denaturant guanadinium hydrochloride (GuHCl) (twice as aggressive), as Tf. Cut II did not unfold in urea. As expected, the results obtained for Tf. Cut II, shows a highly two-state transition unfolding process (midpoint around 2.5 M GuHCl), and matches those seen for other cutinases (Hegde and Dasu, 2014; Otzen, Giehm, ... Pedersen, 2007; Sulaiman, You, ... Kanaya, 2014). Comparably with midpoints of 1.5 M and 3.5 M in urea for ANC Cut-PETase and ANC Cut-Therm respectively, the ancestors have considerably reduced thermodynamic stability (Figure 3.5). The results also suggests that the ANC Cut-PETase is barely stable due to a very short native baseline, and has poorer cooperativity based on the slope of its transition. Thus, while ancestors are folded, they appear to have a significant stability loss in comparison to Tf. Cut II.

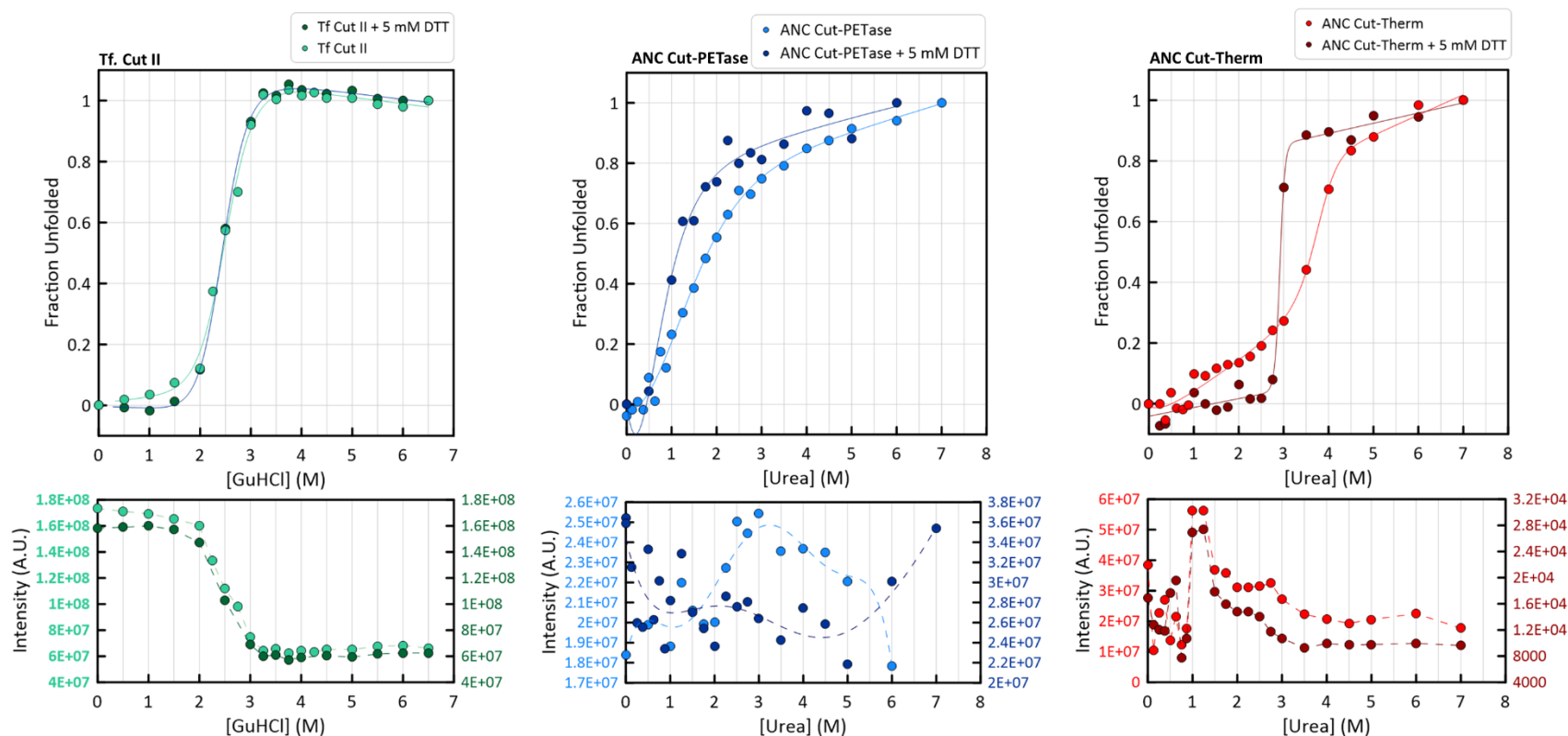


Figure 3.5 Chemical denaturation reveals native baseline for ancestors, but lower thermodynamic stability compared to Tf. Cut II. (Top) Denaturation curves for *Tf. Cut II* (green), ANC Cut-PETase (blue), and ANC Cut-Therm (red) measured in 50 mM sodium phosphate, pH 7.5, 100 mM NaCl at 25 °C following a 30-minute equilibration time. Denaturation curves made in the presence of 5 mM DTT are shaded darker. All curves are fit to Equation 2. Results show high stability for *Tf. Cut II*, compared to ANC Cut-Therm, and the barely stable ANC Cut-PETase. Ancestors show signs of cooperativity, implying the formation of some globular structure. Importantly, this sample for ANC Cut-Therm was shown to contain higher-order disulfide linked samples, and thus, reduced results are likely more representative of the monomers stability. Comparably, the ANC Cut-PETase sample was mainly monomeric, and thus the loss of the disulfide resulted in greater destabilization. The *Tf. Cut II* disulfide bond was not reduced by 5 mM DTT, as shown by iodo-acetamide labelling in APPENDIX I. **(Bottom)** Fluorescence intensity corresponding to denaturation curves. Two axes are shown per graph such that samples with DTT have axis on the right. This is partly due to the large loss in intensity for the

reduced ANC Cut-Therm, compared to its oxidized counterpart. Importantly, the denaturation curve for Tf. Cut II was calculated using intensity data as it fits well into a two-state model. Intensity measurements for the ancestors are more complex, and may correlated to single Trp microenvironment changes, unfolding intermediates, or aggregates. Dotted lines are meant to aid in visualization.

A caveat in the results, is that the two-state model was obtained for the ancestors using max emission wavelength (λ_{\max}) rather than fluorescence intensity. Max emission wavelength is not typically used due to a weak correlation back to protein concentration, and difficulty for further thermodynamic calculations (Monsellier and Bedouelle, 2005). As shown in Figure 3.5, unlike Tf. Cut II, the ancestors fluorescence intensity did not correlate with a two-state model. Instead, they show complex behaviour that may be associated with the formation of stable unfolding intermediates, or aggregates that affects fluorescence intensity but not Trp environment (as the max wavelength remains relatively unchanged). Interestingly, this response to urea has been seen in literature for the Tf. Cut II protein before, where despite the λ_{\max} remaining equivalent (as Tf. Cut II does not unfold in urea), the intensity maxima showed an increase between 1-2 M urea, followed by a steady decrease (Hegde and Dasu, 2014). This coincidentally correlates with the intensity plot for ANC Cut-Therm, where a fluorescence intensity maximum occurs at 1-2 M, prior to reaching an intermediary plateau before a complete unfolding at 3.5 M urea (Figure 3.5). In the paper they speculate that the cause is related to microenvironment changes of surface exposed Trp residues rather than denaturation. This could be possible, as the measurement is collectively probing the behaviour of four trp residues simultaneously, two of which are located on a more surface-exposed loop. However, they (Hegde and Dasu, 2014) also observed this behaviour for Tf. Cut II in GdnHCl, which was not seen here. Due to intensity being linearly correlated to protein concentration, this technique is highly sensitive to errors in volumetric measurement. Knowing this, an analytical balance was used to record the mass of protein added, such that the deviance remained under 2.5%. Hence, given a homologous stock solution, the complex behaviour of the ancestors is thought to be a true representation of Trp environment behaviour, rather than measurement inconsistencies. Thus, at this stage the mechanisms determining their intensity

curves is not known, however, λ_{\max} which is not responsive to intensity effects may act as a reasonable basis for assessing stability.

To further probe the ancestor structure, denaturation curves were also measured following a one-hour incubation with 5 mM DTT. The state of the disulfide was further examined by an iodo-acetamide gel to confirm its condition (APPENDIX I). Results suggest that the Tf. Cutinase II disulfide bond shows a resistance to reduction by 5 mM DTT and is not considered reduced. Small differences between the Tf. Cut II denaturation curves, with and without DTT, may thus be due to the destabilizing effects of DTT itself (Figure 3.5) (Sulaiman, You, ... Kanaya, 2014). This should likewise be kept in mind when considering the ancestor results. A majority of the ancestor sample (70-80%) showed iodoacetamide labelling, and are thought to be reduced, implying that the ancestor disulfide bond is less protected. For monomeric ANC Cut-PETase the loss of the disulfide resulted in loss of the native baseline, and an unfolded structure at 0 M urea (Figure 3.5). Importantly, the sample used for ANC Cut-Therm analysis contained higher-order bands, which were greatly diminished by reduction with 5 mM DTT. Hence, the reduced ANC Cut-Therm denaturation curve is a more valid description for denaturation from a single state (rather than an ensemble which may contain structural variety). This modification is quite noticeable when considering the degree of cooperativity between ANC Cut-Therm with and without 5 mM DTT (Figure 3.5). Intriguingly, the reduction of ANC Cut-Therm results in a devastating loss of Trp fluorescence intensity, which may be due to aggregation. Aggregation of ANC Cut-Therm and Tf. Cut II has been visually observed upon reduction at higher concentrations. Thus, these results imply the importance of the disulfide in ancestor stability, and suggests that the disulfide is more prone to reduction than in Tf. Cut II.

3.5 Inferred Structure of Ancestors by CD, and FTIR

Inference of structure for the ancestors was completed by circular dichroism (CD) and Fourier-transform infrared spectroscopy (FTIR). These methods have complimentary specialities for the accurate assessment of the alpha helical (CD) or beta (FTIR) content of a protein. Indeed, Tf. Cut II has CD results very similar to those predicted by DSSP of its PDB structure (Micsonai, Moussong, ... Kardos, 2022) (Figure 3.6, Table 5). Correspondingly, FTIR of Tf. Cut II matches expectations and presents with alpha helices, intra-beta sheets, and b-turns (Figure 3.7).

The monomeric (higher order structures removed by SEC) ancestors are predicted by BeStSel to have alpha-beta structures (Micsonai, Moussong, ... Kardos, 2022). While this does not confirm a Tf. Cut II-like structure, it does suggest that ancestors are, at the very least, forming helices and beta sheets. Indeed, the predicted % of helices ANC Cut-PETase and for ANC Cut-Therm is 23.2 and 24.2 respectively, which is comparable to the 31.2% in Tf. Cut II. However, the expected peak at 190 nm seems weaker in the ancestors. This observation, and the 10% predicted increase in “other” structure (Table 5), suggests that the ancestors have more unordered structure, and their total content of the alpha helix may be lower.

Notably, 5 mM TCEP has no effect on the content of alpha helix predicted for Tf. Cut II CD, but predicts a significant increase in parallel beta-sheet content (Table 5). As the high tension in CD for reduced samples was relatively higher (high tension of 500 by 185 nm in oxidized samples, compared to high tension of 500 by 200 nm in reduced samples), the fitting and predictions for these results can be impacted. High tension is known to create undesirable noise in CD data that may obscure structural analysis such as predicted secondary structure content. Given the results however, there is the possibility that prediction of alpha helices (which can be viewed

by topology features at 210, and 220 nm) remains possible. Thus, similarly to chemical equilibration results ANC Cut-PETase loses native structure upon reduction, as seen by changes to CD spectra. In comparison, the CD spectra for ANC Cut-Therm are relatively unchanged, with some signs of unfolding.

FTIR of ancestors suggest that ANC CUT-PETase has similar key structural features compared to Tf. Cut II: intra-molecular beta strands around $1630\text{ cm}^{-1}/1690\text{ cm}^{-1}$, and helical signal around 1650 cm^{-1} (Figure 3.7). ANC Cut-Therm shows signs of a helical signal around 1650 cm^{-1} , but lacks signal in intra-beta strand regions. Importantly, aggressive aggregation of ANC Cut-Therm may be visualized here by a strong band around 1610 cm^{-1} ; the formation of inter-beta strands is likewise seen for ANC Cut-PETase but is unrepresented in Tf. Cut II. Taken together with CD data, ancestors reveal signs of forming an alpha-beta hydrolase like structure, but may have locally unfolded regions.

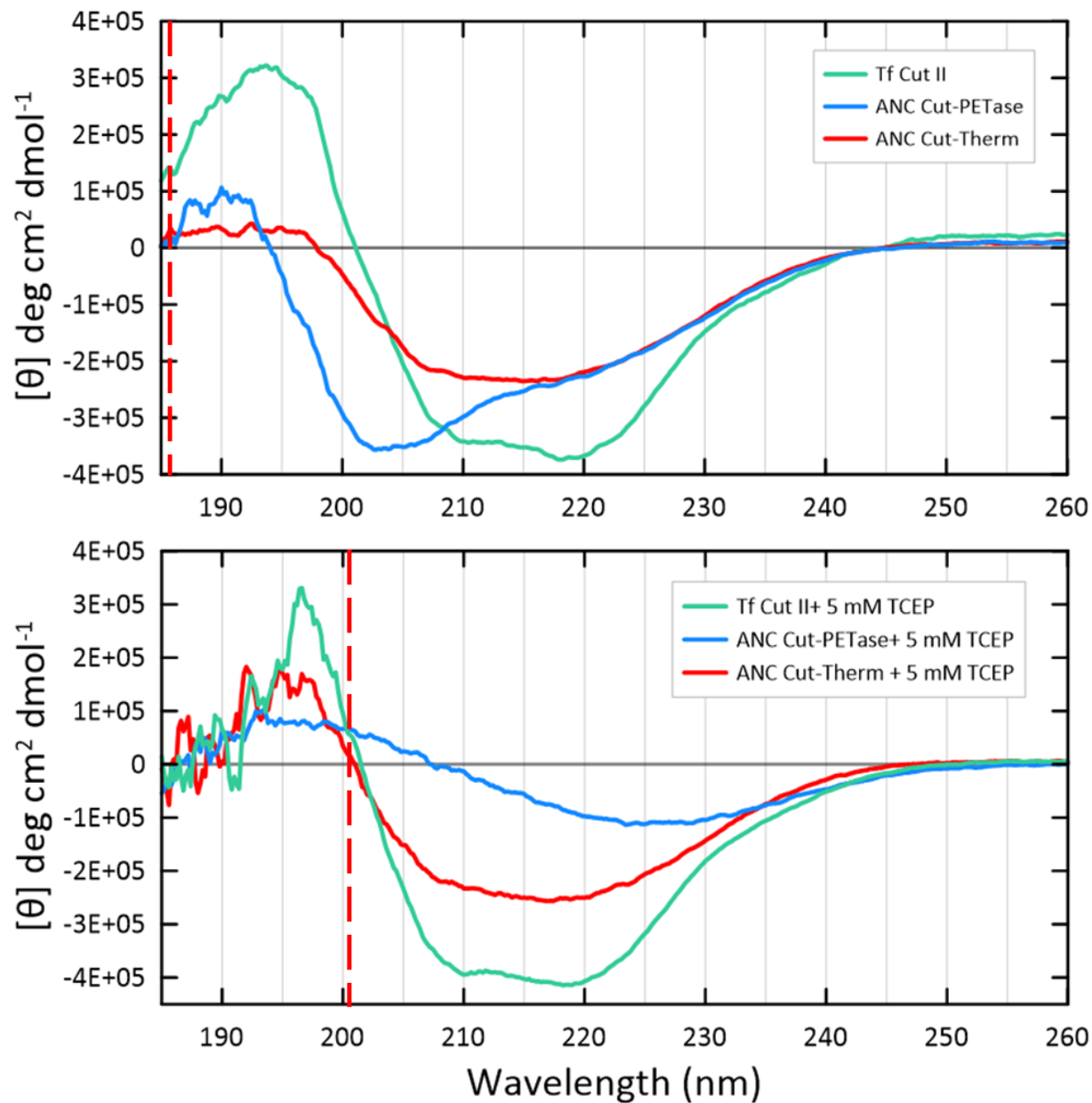


Figure 3.6 CD spectra for Tf. Cut II, ANC Cut-PETase, ANC Cut-Therm. (Top) CD spectra for Tf. Cut II (green), ANC Cut-Therm (red), and ANC Cut-PETase (blue). Red dotted line indicates wavelength at which a high tension of about 500 was reached. **(Bottom)** CD spectra for proteins with addition of 5 mM TCEP for reduction. Dotted line for 500 high tension is shown at 200 nm, which is significantly earlier than for oxidized samples. CD results were obtained by dissolution of protein in water or 10 mM sodium phosphate, pH 7.5 at 25 °C.

Table 5 Predicted secondary structure content from BeStSel. BeStSel predicts the % content of several secondary structures including helices, anti-parallel and parallel beta sheets, turns, and other (which could be perceived to be unordered). The scale factor for amplitude was determined by the BeStSel best factor calculation. Data from the oxidized samples goes accurately down to 195 (400 high voltage), while for the reduced samples the accuracy is significantly less (550 voltage at 200). Finally, the predicted class is based on comparisons to other similar CD spectra.

Method	Helix	Antiparallel	Parallel	Turn	Other	NRMSD	Predicted Class
DSSP from Tf. Cut II	25.5	8.2	12.4	13.5	40.4	-	-
Tf. Cut II Oxidized	31.2	8.6	13.7	8.5	37.9	0.01471	Alpha Beta
Tf. Cut II Reduced	31.4	4.0	28.1	0	36.6	0.10927	Alpha Beta
ANC Cut-PETase Oxidized	23.3	16.1	2.5	0	58.0	0.08245	Beta or Alpha Beta (No structure predicted)
ANC Cut-PETase Reduced	1.0	0	33.9	19.2	46	0.06326	Mainly Beta
ANC Cut-Therm Oxidized	24.2	15.4	11.8	4.5	44.1	0.06594	Alpha Beta
ANC Cut-Therm Reduced	27.3	5.7	23.0	0	43.9	0.07525	Alpha Beta (no structure predicted)

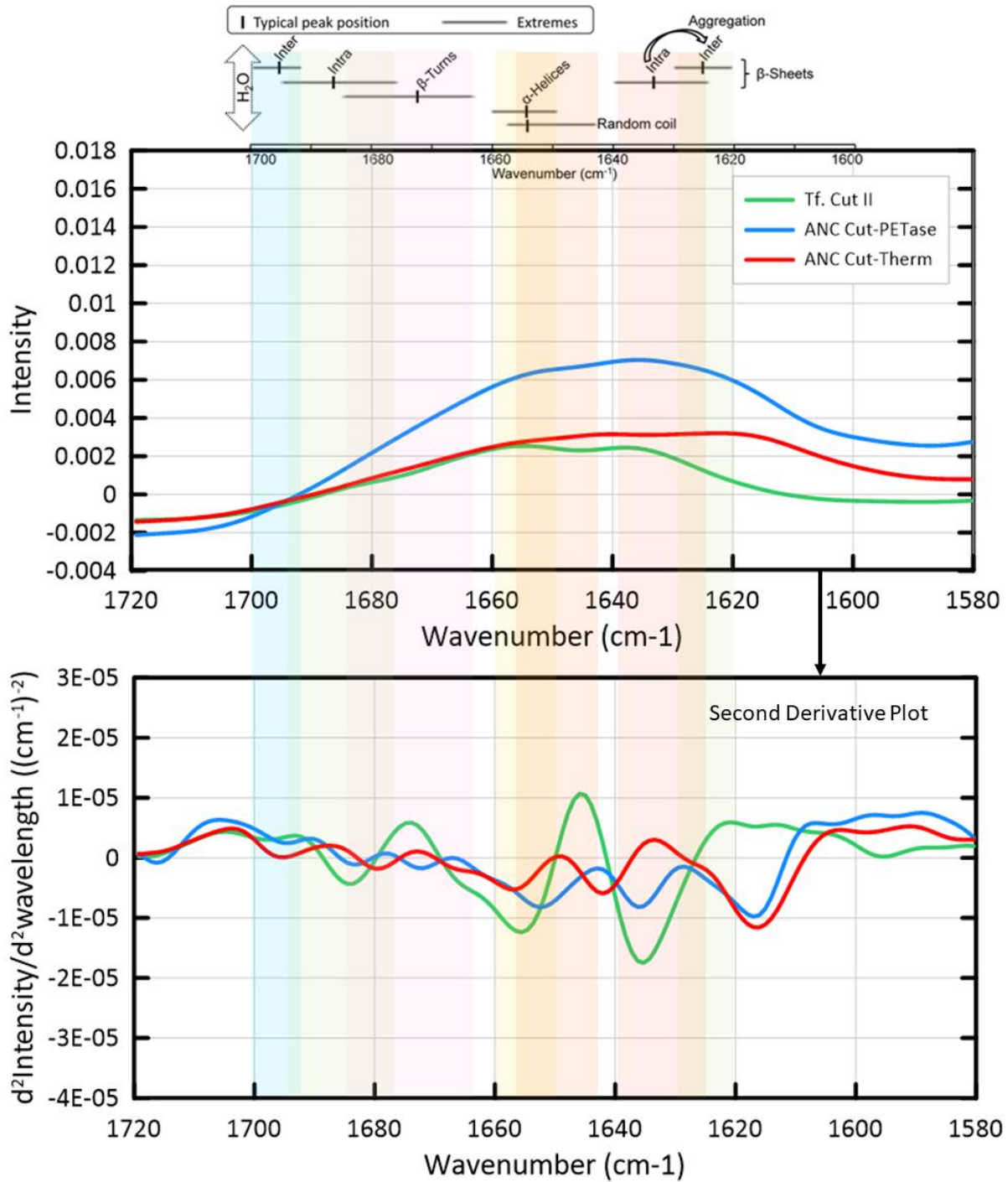


Figure 3.7 FTIR spectra for Tf. Cut II, ANC Cut-PETase, ANC Cut-Therm. **(Top Right)** FTIR spectra for proteins in amide I region. Expected assignments for proteins in water are shown as coloured regions, with the legend being obtained from (Natalello and Doglia 2015). **(Bottom Right)** First derivative plot for FTIR spectra. FTIR results were obtained in 50 mM ammonium carbonate for ancestors, and 50 mM Tris, pH 8 for Tf. Cut II at 25 °C.

3.6 Enzymatic Activity of Tf. Cut II

Enzymatic assays were performed for Tf. Cut II with a model substrate, p-nitrophenol acetate (pNPA). Esterases are typically able to hydrolyze this substrate into acetate and p-nitrophenol, the latter of which is observable by absorbance at 405 nm. Importantly, the substrate is not miscible in water, and thus dilution into some solvent is necessary (i.e., acetonitrile, isopropanol, ethanol). Thus, the effect of increasing ethanol, dimethyl sulfoxide (DMSO), and acetonitrile percentage on enzyme activity was measured (Figure 3.8). DMSO showed very little effect to enzyme activity, while ethanol and acetonitrile linearly improved and impeded activity respectively. For this reason, the use of DMSO may have been optimal, however, a yellow colour (indicating the presence of the product, p-nitrophenol) soon appeared following dissolution of substrate into DMSO suggesting the presence of a non-enzymatic reaction. Consequently, ethanol was used at 2.5% to match as closely as possible to the perhaps unaltered activity in DMSO.

Hydrolysis kinetics were performed with varying substrate concentration for Tf. Cut II, revealing a K_m of 1.1 mM and V_{max} of 1.0 mM/min (Figure 3.8). Further, residual pNPA activity following a 30-minute incubation time at varying temperature revealed losses of activity beginning at 67 °C (~44%), with complete depletion at 70 °C. These correlate well with temperature denaturation experiments for Tf. Cut II, and to similar experiments done previously (APPENDIX H) (Roth, Wei, ... Sträter, 2014). It was desired that these results could be compared to ancestor activity, however, ancestor activity could not be detected above the blank pNPA hydrolysis rate even using concentrated protein (a max concentration of 4-5 mg/mL) (Figure 3.8). Attempts were initially made on dilute ancestors (for reference Tf. Cut II activity assays are completed around 0.1 ug/mL), however under no concentration conditions could activity be confirmed. Thinking that perhaps very minute signals for activity were being hidden in the blank hydrolysis rate, another

substrate pNPP was tried. Unlike, pNPA, pNPP is typically insoluble and has an insignificant background hydrolysis rate. Thus, enzyme activity can be confirmed simply by observing a yellow product. However, even given sufficient time—observed pNPP wells for hours—there was no alteration to color (Tf. Cut II well turns yellow immediately) (Figure 3.8). Given the degree of structure loss observed in CD and FTIR (APPENDIX J), this is perhaps not an unexpected outcome. The active site relies on the proper conjunction of three loops (connecting beta strands to alpha helices) in order to form the catalytic circuit. It is likely the intersection could be destabilized.

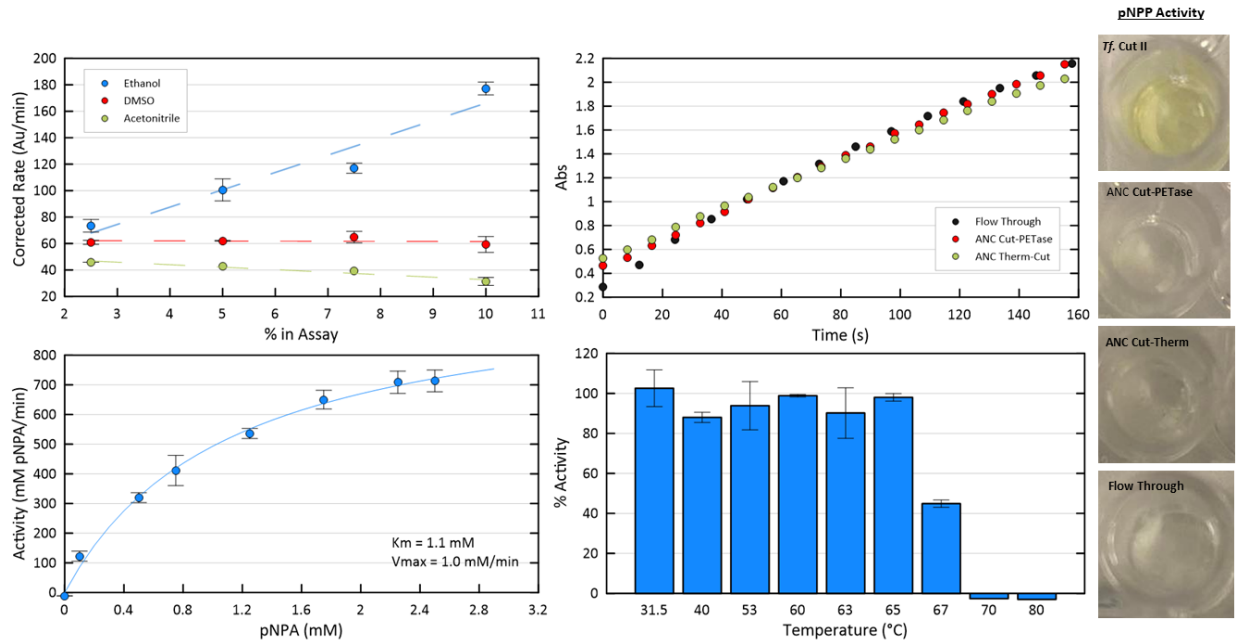


Figure 3.8 Hydrolysis of pNPA with Tf. Cut II. (Top Left) Effect of solvent choice on enzymatic activity for pNPA dissolution. Ethanol (blue) shows a positive effect on activity with increasing concentration. Dissimilarly, acetonitrile causes a slight decrease in activity with increasing concentration. While, DMSO has no effect on enzyme activity, it presented as a yellow coloured solution, suggesting higher levels of non-enzymatic hydrolysis compared to its counterparts. Thus, 2.5% ethanol was used to measure pNPA hydrolysis. (Top Right) Background hydrolysis rate of pNPA with ancestor AMICON flow-through (blank, 50 mM ammonium carbonate, pH 8.9, black) compared to pNPA incubated with concentrated ancestral proteins (~4.3 mg/mL, red/yellow). This reveals a lack of ancestor activity above baseline. (Bottom Left) Michaelis-menton kinetics for Tf. Cut II and pNPA reveals K_m of 1.1 mM and V_{max} of 1.0 mM/min. All measurements were completed in triplicate in 50 mM sodium phosphate, pH 7.5, 100 mM NaCl at 25 °C, and p-nitrophenol produced was measured at 410 nm. (Bottom Right) Residual activity was determined after 30 minute incubation of Tf. Cut II. Then pNPA activity was measured and compared to activity at 25 °C (set to 100%). Measurements were completed in 50 mM sodium phosphate, pH 7.5, 100 mM NaCl at 25 °C, and p-nitrophenol produced was measured at 410 nm. (Right Side Panel) Incubation of Tf. Cut II (top), ANC Cut-PETase (2nd), ANC Cut-Therm (3rd), and flow through (bottom, blank) with 0.7 mM pNPP. Lack of yellow colour in ancestral samples, suggest that ancestors were not capable of digesting pNPP.

3.7 Sequence and structural based analysis of ancestors

3.7.1 *Building of homology models, and overall predicted variation*

Computational approaches for protein structure prediction are either template based (relying on homologous structures and sequences), or template-free (ab-initio) modelling. Given several close sequence homologs available for the ancestors, and the limitations for Robetta ab-initio to properly assume the structure of Tf. Cut II (too large for prediction), ancestor models were created using template based methods (Bonneau, Tsai, ... Baker, 2001). Accordingly, SWISS-Prot, Robetta CM, Robetta TTA, Phyre2, I-TASSER, and Alpha Fold2 predict global structures for ancestors that highly relate to PETase structural homologs ($< 2 \text{ \AA} \text{ C}\alpha \text{ RMSF}$) (APPENDIX J) (Baek, DiMaio, ... Baker, 2021; Bonneau, Tsai, ... Baker, 2001; Jumper, Evans, ... Hassabis, 2021; Kelley, Mezulis, ... Sternberg, 2015; Waterhouse, Bertoni, ... Schwede, 2018). In order to choose a model for further analysis, models were first compared to one another by their predicted FoldX and Rosetta energy scores (Alford, Leaver-Fay, ... Gray, 2017; Park, Bradley, ... DiMaio, 2016; Schymkowitz, Borg, ... Serrano, 2005; Williams, Headd, ... Richardson, 2018). However, results suggested that predictors are biased to structures created by certain programs, likely those prioritizing similar energy functions—for example, FoldX strongly preferred models created by Phyre2, while Rosetta strongly favoured Robetta TTA/CM models. Thus, in lieu of stability predictions, MolProbity Ramachandran scores of models was used for their assessment—particularly to weed out models that predicted a large numbers of outliers or unideal geometries (i.e. twisted peptide chains) (Williams, Headd, ... Richardson, 2018). This analysis showed poor overall scores for I-Tasser and an increased number of Ramachandran outliers for Phyre2. The remaining models had comparable Ramachandran scores, thus, rationally choosing one over the other became difficult (APPENDIX K). Ultimately, at this stage, the alpha fold models chosen for

further analysis (molecular dynamics, structure-based predictions) due to the method's superior performance at CASP14 (Pereira, Simpkin, ... Lupas, 2021).

The majority of disparities between all models are found in loops, particularly at the sites with three-residue loop insertions and the C-terminal (APPENDIX J). As a reminder, at these locations the ancestor loop lengths were fixed to match Is. PETase (thus are 3 residues longer than Tf. Cut II at loop #1 and #2) or Tf. Cut II (C-terminal). From detailed observation of structures of extant proteins with both 3-residue loop insertions, it seems that the inserts are partly responsible for extending upstream alpha helices as well as extending the loop itself. This in turn modifies the behaviour of the downstream loop, even when sequences remain similar between enzymes with both loop inserts vs. with neither one (now called INSEQs and GAPSEQs respectively for succinctness—see APPENDIX D for a visual of these groups). Further, following analysis of INSEQs in the alignment, it is clear that ancestral sequence reconstruction, which treats amino acid sites as independent, may not be capable of detecting interdependent relationships when alignment coverage of sites is low or distantly located on the phylogenetic tree. In the following section, residue variation that seem to be co-dependent on insert existence are explored.

To investigate insert based permutations at the site of insert loop #1, sequences were split between those with (INSEQs) and without the loop insert (GAPSEQs) (Figure 3.9). This separation revealed a conserved arginine (R120, R122, R127) trio in GAPSEQs. R122 and R127 seem to be implicated in the stabilization and linking of loop #1 to neighbouring loops in Tf. Cut II (GAPSEQ). From observations of the Tf. Cut II structure, arginine residues seem to be commonly located at interfaces, and could be acting as a structural *glue* for loop regions. Additionally, they seem to be disproportionally present (12:4) on the C/N terminal hemisphere, opposite of the active site. The final arginine in this particular trio, R127 (rarely K127), is conserved in all sequences,

and is likewise surrounded by highly conserved residues (such as D124), suggesting that it is an important stabilizing element. Notably, GAPSEQs all possess a small helical turn that redirects the loop back toward the main structure (Figure 3.9). When the insert is present, this helical turn structure is lacking due to a significant gain of a proline residue at position 118. The role of this proline may be twofold: (1) if the upstream helix extension is sterically too close to the helix turn, the proline could break helix turn formation and reduce tension, and (2) kinking of the loop by the proline in this area may help with redirecting the loop toward the main structure.

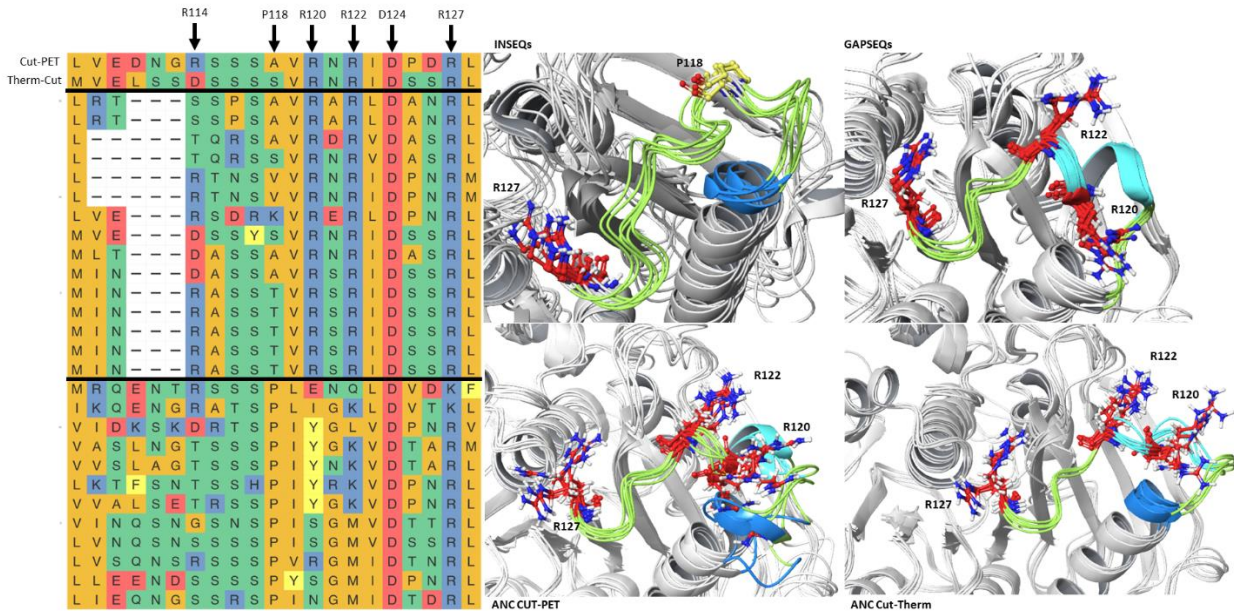


Figure 3.9 Presence of loop insert #1 is coupled with gain of P118 and loss of conserved Arg trio. (Left) Multiple sequence alignment with sequences for ANC Cut-PETase and Cut-Therm at top, followed by GAPSEQs in the centre, and INSEQs at the bottom. Notable residues are highlighted by a black arrow. Specifically, these are P118 (exclusively found in INSEQ sequences), R120/R122 (conserved in GAPSEQ sequences), and D124/R127 or K127 (interacting residues conserved in all sequences). (Right Top) Crystal structures of INSEQs (top left), and GAPSEQs (top right). Loop #1 is shown in light green, insert is shown in dark blue (for INSEQs), and helix turn is shown in light blue (for GAPSEQs). For GAPSEQs, conserved arginine residues are also shown with R127 the furthest left, followed by R122 and R120. Notably, the conserved R122 and R120 are not present in INSEQs. Instead, INSEQs have a conserved P118, which is shown by yellow sticks. (Right Bottom) Model structures of ANC Cut-PETase (bottom left) and ANC Cut-Therm (bottom) created by Alpha Fold, Robetta CM, Robetta TTA, Phyre, Swiss-Prot, and I-TASSER. Models for ANC Cut-PET show variation in helix propensity of loop insert (dark blue), and presence of helix turn (light blue). Meanwhile, models for ANC Cut-Therm agree more closely, and predict a helix extension followed by loop. Notably, models for both ancestors occasionally predict an altered R127 rotamer.

ANC Cut-Therm is predicted by almost all models to have an extended helix structure as seen in INSEQs. However, its sequence does not contain P118, and retains the entirety of the R-trio, which is an unrepresented combination and may be unfavourable. Interestingly, for ANC Cut-PETase, there seems to be a lack of consensus by models for the helix extension propensity of its Loop 1 insert. In fact, models show signs of loop confusion, where it seems ANC Cut-PETase is obtaining blueprints for both INSEQ and GAPSEQ (i.e., presence of helix turn, vs. not). The

possibility of helix extension coupled with a non-zero probability of helix turn formation may ultimately result in loop structures and conformations that are not reminiscent of extant proteins for both ancestors. This instability may be informally observed by several models (particularly Robetta TTA, CM, Phyre) non-trivially predicting an altered rotamer for R127. Thus, using insert-based perturbations of amino acid presence in the multiple sequence alignment, it is perhaps possible to justify that the inclusion of the insert at loop #1, coupled with retainment of the R-network (R120, R122), as well as a lack of a proline residue, resulted in increased loop flexibility in the ancestors, which may maintain the formation of important conserved interactions.

The second insert is predicted to extend an alpha helix/loop leading up to the cHis residue. Inspection of the loop reveals that the majority of predicted interactions between the cHis loop and surrounding chains is on the whole similar between Tf. Cut II (GAPSEQ) and Is. PETase (INSEQ). The very notable exception is an insert-dependent addition of a disulfide bond between C177-C214 that is responsible for anchoring the cHis and cAsp loops in INSEQs (exceptions are PET 27 and PET 30). Active site disulfide knockouts in Is. PETase have previously been shown to inactivate the enzyme (Fecker, Galaz-Davison, ... Ramírez-Sarmiento, 2018; Joo, Cho, ... Kim, 2018). Investigation of Is. PETase (INSEQ) with a *reduced* active site disulfide by molecular dynamics reveals increased fluctuation, and increased variability in loop behaviour compared to the intact disulfide model (Figure 3.10). This is coupled with an increase in average distance between cHis (ϵ N) and cSer (O) from 3.14 Å to 4.01 Å. Similar trends may be apparent in ancestors, particularly ANC Cut-Therm, which experience increased loop fluctuation and cHis-cSer distances—ANC Cut-PETase average distance is 3.7 Å, ANC Cut-Therm average distance is 5.2 Å (APPENDIX L). This said, distances were not improved by removal of loops in ANC Cut-PETase (3.9 Å average), or addition of disulfide (3.9 Å average) (APPENDIX L). In ANC Cut-

Therm, apart from the 3 residue-insert, the cHIS loop has a sequence identical to that in Thh_Est. Thus, as expected, following loop removal the RMSF for the cHis loop in ANC Cut-Therm was reduced and the average distance between cSer and cHis was lowered to 4.9 Å. Importantly, molecular dynamics analysis of this area in wildtypes often gives diverse results: for example, cSer134 can interact with nearby H133 in GAPSEQs which increases the perceived average distance between cSer and cHis (this is seen in ancestors), and even the Is. PETase cHis loop has been seen to delocalize from the active site. Thus, it may be more difficult to make predictions on this area solely based on simulations, and perhaps greater evidence of disulfide need is obtained colloquially from Is. PETase experimental disulfide knockouts.

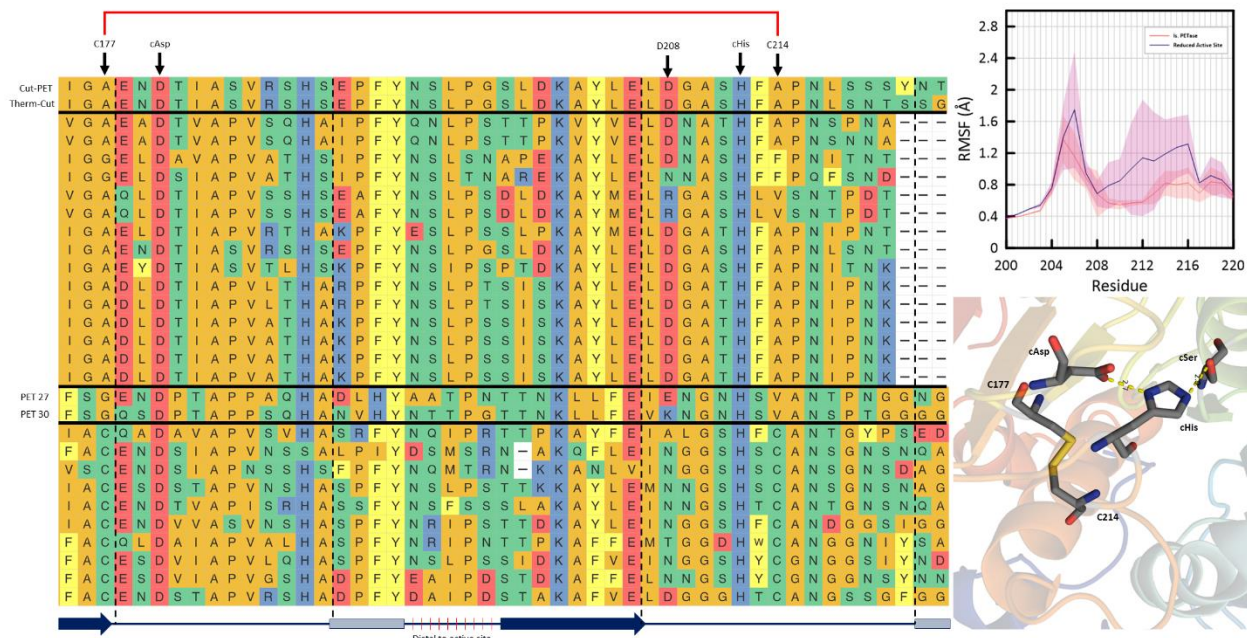


Figure 3.10 Loop insert #2 is implicated in active site coordination. (Left) Multiple sequence alignment of active site region containing loop insert #2. Importantly, secondary structure elements are traced along the bottom of the alignment, with dotted lines acting as an extension to show loop sections. Two catalytic residues, cASP and cHis are annotated by a black arrow. Top two sequences are ANC Cut-PETase and ANC Cut-Therm. These are followed by GAPSEQs, PET27/PET30, and INSEQs. Upon insert addition, a notable addition of a disulfide bond between C177 and C214 are observed (exception is PET27/PET30). **(Right Top)** Average RMSF (n=3) of Is. PETase with intact disulfide bond (red), and reduced disulfide bond (purple), zoomed in on cHis loop. On reduction, average RMSF is increased, and variability of conformations is increased (shown by increased standard deviation between simulation repeats). **(Right Bottom)** Structure of Is. PETase (PDB 6EQE) showing catalytic triad, and disulfide.

The disulfide is not required for GAPSEQ enzymes, and is shown only in insert containing loops—a link that was not made by ASR or by visualization of the alignment due to limited representation. The alliance between catalytic residues is required for activity, and it is likely that the increased movement of the loop is then detrimental to enzyme productivity. The rare case of PET 27 and PET 30, which have an extended loop, but no active site disulfide remains to be explained. By studying the crystal structure of PET30 (PDB 7PZJ) however, one could anticipate

that the active site remains stabilized due to increased proline content in the cAsp loop and cHis loop (compared to other INSEQs), and additional interaction possibilities from seemingly closer beta strands. This said, PET 27 and PET 30 are quite distinctive from other PET hydrolases in sequence and have several distinct features (such as an additional N-terminal disulfide). Thus, while it is certainly possible to stabilize the cHis-cAsp connection in insert containing sequences without the disulfide bond, it is probable that like Is. PETase disulfide knockouts, the ancestors were not in a situation to do so.

3.7.2 *Sequence analysis of ancestors compared to known PET hydrolases*

In the previous section, the role of additional loop inserts was investigated, revealing several sites of loop related amino acid perturbations. Here, the remainder of the sequence is analyzed for possible site mutations that may have also contributed to the reduced stability of the ancestral proteins. In order to provide a frame of reference in the vast number of sequence homologs available for comparison, the closest extant sequence for both ancestors was sought. Perhaps sensibly, this search revealed Thh Est, which is the nearest descendent for both ancestral nodes, to be the closest match (84% identity to ANC Cut-PETase, 93% identity to ANC Cut-Therm). Thh_Est is a characterized esterase from *Thermobifida halotolerans* that has a low pI (5.24), soluble expression in *E. coli*, and activity against PET polymers (Ribitsch, Herrero Acero, ... Guebitz, 2012). Its structure, thermostability, and halotolerance are as of yet uncharacterized, but due to its host ability to grow in conditions up to 50 °C (similar to others from the Thermobifida family), and in up to 10% (w/v) NaCl, it is likely that Thh_Est, has some tendency for these properties (Yang, Tang, ... Li, 2008). Thus, assuming this as an existing stable reference point,

ANC Therm Cut has a total of 11 site mutations from Thh_Est, while ANC Cut-PETase has a total of 38 (excluding loops in both cases).

FoldX position scan was used to predict changes to stability ($\Delta\Delta G$) to Thh_Est upon point-mutation to the ancestral residue (Table 6) (Schymkowitz, Borg, ... Serrano, 2005). Of the 11 site mutations in Cut-Therm, 10 match consensus residues for known GAPSEQs—the remaining one is a surface Ser to Ala mutation that does not seem to alter any valuable interactions. Further, none of these perturbations are predicted to have significant destabilizing $\Delta\Delta G$ values, except for F256, which could be considered destabilizing in Thh_Est due to steric hinderance with the neighbouring W255 residue (L255 in ANC Cut-Therm). Thus, it may be possible that the majority of activity and stability issues in the Cut-Therm ancestor are based on loop insertions.

ANC Cut-PETase has 38 site mutations from Thh_Est, making it quite complicated to unravel one single cause for instability. Unlike for ANC Cut-Therm, many of these mutations are clustered together, and combinatory effects are quite possible. However, using consensus matching and $\Delta\Delta G$ s from Thh_Est as a guide, particularly destabilizing residues may be hypothesized (Table 6). Upon initial analysis, it was observed that 17 of the 38 residues match GAPSEQ or INSEQ consensus: these residues may be less likely to impair structure. Second, most mutations seem to cause small changes to the predicted $\Delta\Delta G$ (considered -1 to 1). Of the mutations considered destabilizing ($>1 \Delta\Delta G$), matching environments containing the varied residue are found in related PET hydrolase such as Cut 190 and PSH1—these include T47, Q78, G98, D126, F83, and A143. This may indicate that these changes are tolerable when taken in consideration with other nearby changes. However as some of these are considered substantially destabilizing and are buried (F83, Q78, G98, A143), they should be kept in mind as their combined effect could result in unforeseen

destabilization. On top of these changes, the ANC Cut-PETase also contains loop inserts, which is likely a major destabilizing factor in ANC Cut-Therm.

Table 6 Changes to ANC Cut-Therm and Cut-PETase sequence with respect to closet sequence match, Thh_Est. Table shows residues that have changed with respect to Thh_Est (11 for ANC Cut-Therm, 38 for ANC Cut-PETase). Blue highlights represent residues matching consensus of GAPSEQ, INSEQ or both. $\Delta\Delta G$ values were calculated by Fold-X position scan and are based on changing Thh_Est residue to ancestor residue; green highlights correspond to stabilizing changes (< -1), while red corresponds to destabilizing changes (> 1). For values of interest, notes about structural placement were considered.

Residue #	ANC Cut-Therm					$\Delta\Delta G$	Additional Notes
	ANC	Thh_Est	GAPSEQ Conserved	INSEQ Conserved			
11	E	N	D	V	-1.32	Forms interaction to K227 on nearby loop	
22	S	R	S	S	0.72		
32	A	Q	S	S	0.49		
34	S	R	S	S	0.17		
49-50	EG	--	QN	TG	--		
111-113	LSS	---	---	QNG	---		
117	S	Y	S	S	0.22		
130	V	A	V	V	-1.71	Buried valine	
151	K	Q	K	K	-0.39		
161	L	T	L	S	-0.72		
166	S	G	S	S	-0.46		
221-223	SSG	---	---	YND	---		
255	L	W	L	L	0.08		
256	F	G	F	S	3.43	Might be destabilizing in Thh_Est due to W255 (L255 in ANC)	
ANC Cut-PETase							
8	D	N	N	D	0.19		
11	E	N	D	V	-1.32	Forms interaction to K227 on nearby loop	
22	S	R	S	S	0.72		
27	S	R	N	N	0.80		
30	S	R	R	S	0.75		
32	A	Q	S	S	0.49	Remodelling of loop in entirety. May have loss of interaction to nearby helix.	
33	V	A	A	V	6.95		
34	S	R	S	S	0.17		
47	T	N	N	A	2.10	Alters loop packing. Identical to Cut 190.	
49-50	EG	--	QN	TG	--		
58	A	S	S	V	0.03		
63	A	G	G	A	-0.71		
64	S	T	T	A	0.09		
68	M	I	I	I	0.59		
78	Q	H	H	H	2.32	Quite buried, may cause problems. Matches environment in PSH1.	
83	F	M	I	I	5.83	Buried Phe. Next to second Phe, Val, Leu, Ile. Might not be space for accommodation. Matches Cut 190 environment, but displacement	

						of F36 due to changes in 29-38 loop causes loss of FF interaction.
90	R	T	T	G	-2.02	
98	G	A	A	A	1.59	Technically loses hydrophobic Ala interactions, but this Gly is supported in Cut190, PH7, PSH1
99	D	S	R	S	0.55	
102	S	D	N	M	-0.14	
108	L	M	M	L	0.11	
111-113	DNG	---	---	QNG	---	
117	S	Y	S	S	0.22	
118	A	S	A	P	-0.33	
125	P	S	S	T	-1.24	Increased loop rigidity inferred by Pro. In loop with insert #1.
126	D	S	S	N	1.40	D124 is nearby and has a conserved reaction with R127. Perhaps the in-between D126 causes a charge clash. Present in Ple629, PET27.
130	V	A	V	V	-1.71	Buried valine.
142	A	R	R	R	0.71	
143	A	L	L	V	3.11	Loss of contacts made by Leu. Similar to environment in PSH1 – may be tolerated.
146	D	R	Q	D	0.72	
149	S	D	D	R	0.17	
151	K	Q	K	K	-0.39	
161	L	T	L	S	-0.72	
166	S	G	S	S	-0.46	
167	E	S	S	S	0.24	
221-223	YNT	---	---	YND	--	
244	S	T	T	S	0.55	
255	L	W	L	L	0.08	

Finally, to conceptualize how applicable this ASR attempt is for improving PET hydrolase activity, both ancestors are compared to known activating and stabilizing literature point mutations (Figure 3.11) (APPENDIX A). A very important caveat to consider in this comparison, is that all literature point mutation considered here were completed on different enzymes, and thus, variations in sequence of the non-mutated structures are not accounted for. Hence, a residue that increases activity in one enzyme, does not ensure increased activity in another. Further, reports of activity and thermostability are often interwoven for PET hydrolases due to untypical changes to the substrate morphology—a PET hydrolase with increased thermostability may be perceived as more active at high temperatures, but actually have less activity when measured at low

temperatures. In extension, varying substrate crystallinity, and testing conditions (pH, salts) can impact the representation of activity. Nonetheless, ANC Cut-Therm, presents a gain of five activating, eleven deactivating, eight stabilizing, and eight destabilizing mutations. Very similarly, ANC Therm-PET has four activating, twelve deactivating, nine stabilizing, and seven destabilizing mutations (Figure 3.11). Thus, results do not seem too convincing for theoretical improved PET hydrolase activity based on mutations that are already known. Interestingly, seventeen of the twenty-five activating mutations are found in sequences for PET hydrolase enzymes listed in Table 1—particularly among INSEQs. This realization suggests that if choice of extant sequences in the alignment was restricted to this smaller more active subset, it could boost the opportunities of accruing activity improving residues.

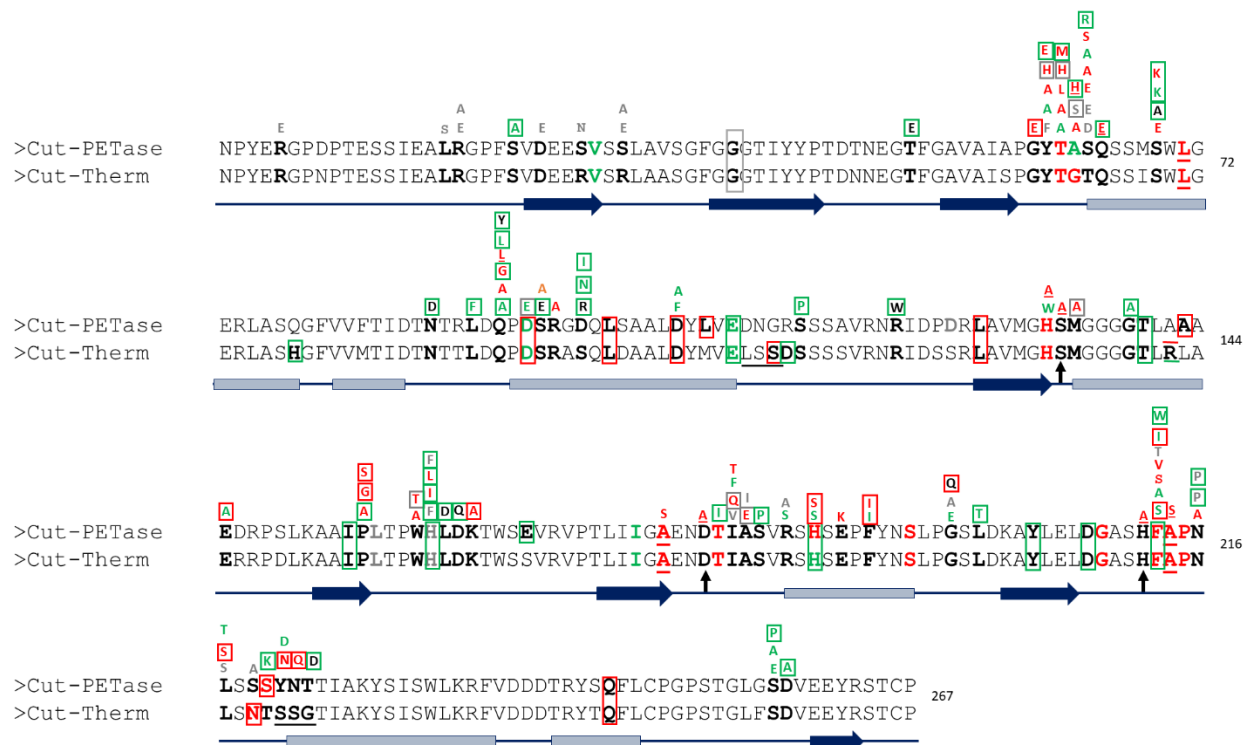


Figure 3.11 Comparison of ancestor sequences to literature mutations previously tested on a variety of PET hydrolase enzymes. All bolded residues are locations of tested sites. Tested literature residues are written over the residue of the site that was altered. Residues colored red or green represent a loss or gain in activity. Similarly, residues outlined in red or green, represent a loss or gain in stability. Finally, secondary structure of ANCs is directly below the sequence: helices are represented by a light blue box, beta strands are represented by a dark blue arrow, and loops are shown as lines.

3.7.3 Analysis based on stability and solubility predictors

Numerous stability and solubility predictors have been constructed in order to give protein engineers an advantage when considering new designs—sometimes to predict disastrous outcomes, or to find property ameliorating mutations. Several of these were used in order to determine if these were capable of predicting the poor solubility, aggregation liability, and reduced stability of the ancestors. They range from long-present stability predictors (instability index, aliphatic index), structural complexity predictors (LRO, ACO), and aggregation predictors (AGGRESCAN, CAMSOL) (Broom, 2016; Conchillo-Solé, de Groot, ... Ventura, 2007; Gromiha and Selvaraj, 2001; Guruprasad, Reddy, and Pandit, 1990; Ikai, 1980; Ivankov, Garbuzynskiy, ... Finkelstein, 2003; Sormanni, Aprile, and Vendruscolo, 2015). Results reveal that predicted ancestor values for all strategies lie within values for other known PET hydrolases, and they are even considered less prone to aggregation by AGGRESCAN and Camsol (Table 7). Thus, these results suggest that there is nothing alarming that is predicted here that may be an indicator to the reduced stability, and folding seen in ancestors.

Table 7 Predictions of stability and solubility by ProtParam, LRO/ACO, and AGGRESCAN/Camsol. Ancestor properties are compared to properties of 4 known PET-hydrolases Tf. Cut II, LCC, Is. PETase, and Thh_Est. Properties compared are protparam predictions (Instability Index, Aliphatic index), predictors of long-range and absolute contact orders (LRO/ACO), and solubility (AGGRESCAN, Camsol). Results considered the ‘worst’ of the group are colored red, while best results are colored ‘green’. Ancestors are typically within, or above values predicted for other PET hydrolases.

Method	Tf. Cut II	LCC	Is. PETase	Thh_Est	ANC Cut-PETase	ANC Cut-Therm
Instability Index	37.29	34.22	42.10	40.11	39.26	41.09
Aliphatic Index	80.80	77.47	62.53	68.55	71.65	72.36
LRO	6.90	7.17	7.02	7.02	6.98	6.91
ACO	24.32	25.08	25.50	24.70	25.46	24.90
AGGRESCAN (a ³ v)	-0.032	-0.009	-0.059	-0.102	-0.1	-0.089
Camsol Structurally Corrected	1.91	1.74	1.84	1.10	2.08	1.99

3.7.4 Disulfide bond formation

Ancestors experimentally present with C-terminal disulfide formation difficulty, and routinely show inter-molecule disulfide linkages. In structural homologs this disulfide is conserved and responsible for linking an extended loop region to the final beta strand of the enzyme. Knockouts to this disulfide in LCC have shown reduced stability (Sulaiman, You, ... Kanaya, 2014). In the original alignments, the length of the extended loop between the Cys residue and the start of the final beta strand is quite variable, and ranges from 7-11 amino acids long (with loops of 8 residues being the most common). Many sequences originating from thermotolerant bacteria have loops of 11 amino acids in length, and thus loop length was set at 11 for the ancestors. Importantly, this span of residues predicted by ASR for ANC Cut-PETase is identical to those in two other extant sequences, whereas ANC Cut-Therm has a unique loop sequence.

Molecular dynamics for this region, similar to the N-terminal region, typically shows high fluctuation due to its general lack of structure. Curiously, Is. PETase has significantly reduced loop RMSF in its terminal (8 residues long), compared to Tf. Cut II (11 residues long), perhaps due to its shorter loop length. Ancestors both reveal RMSF values in this region similar to those seen for Tf. Cut II. These results, however, are based on an already pre-formed disulfide bond. Thus, to further inspect the feasibility of disulfide formation, reduced disulfide simulations were conducted. Notably, in reduced simulations, there is an assumption that the protein has folded prior to forming disulfide links. This may be reminiscent of how folding was attempted in the lab, where DTT was added during urea dialysis in order to promote folding prior to linking. Order is incredibly important as early disulfide links can significantly alter the folding landscape (Muttathukattil, Singh, and Reddy, 2019). On the other hand, any misfolding that dislocates the

position of the Cys loop, will likely prevent disulfide formation and allow opportunities for inter-linking.

Results of reduced simulations reveal an ultimately unchanged RMSF pattern for the overall loop. However, significant changes to average cysteine distance were identified (Figure 3.12). In the reduced form, the average distance of Tf. Cut II and Is. PETase cysteines hovers around 4.4 and 4.6 Å. Meanwhile distances for ancestors are increased to 6.7 and 7.4 for ANC Cut-PETase and Cut-Therm respectively. For ANC Cut-PETase this is not due to massive rearrangements, but rather due to a rotamer change, where the loop CYS opts for interaction with a backbone oxygen, and the terminal CYS is turned away towards the solvent. Contrarily, ANC Cut-Therm experiences a *slipping* of the loop CYS away from the terminal CYS, that is seemingly caused by reorganization of the loop turn and its subsequent movement toward the N-terminal loop (F256 and A16 interact). These results suggest that the ANC Cut-Therm C-terminal may be stabilized in a way that obstructs disulfide formation (assuming disulfides are formed after folding of this region). As well, from RMSF discrepancies between Is. PETase and Tf. Cut II in the terminal loop region, it is likely that a shorter loop may have been easier for ASR to accurately stabilize. This is in part due to the non-trivial need for Tf. Cut II to retain proper orientation of the disulfide despite one CYS being present in a highly flexible loop region.

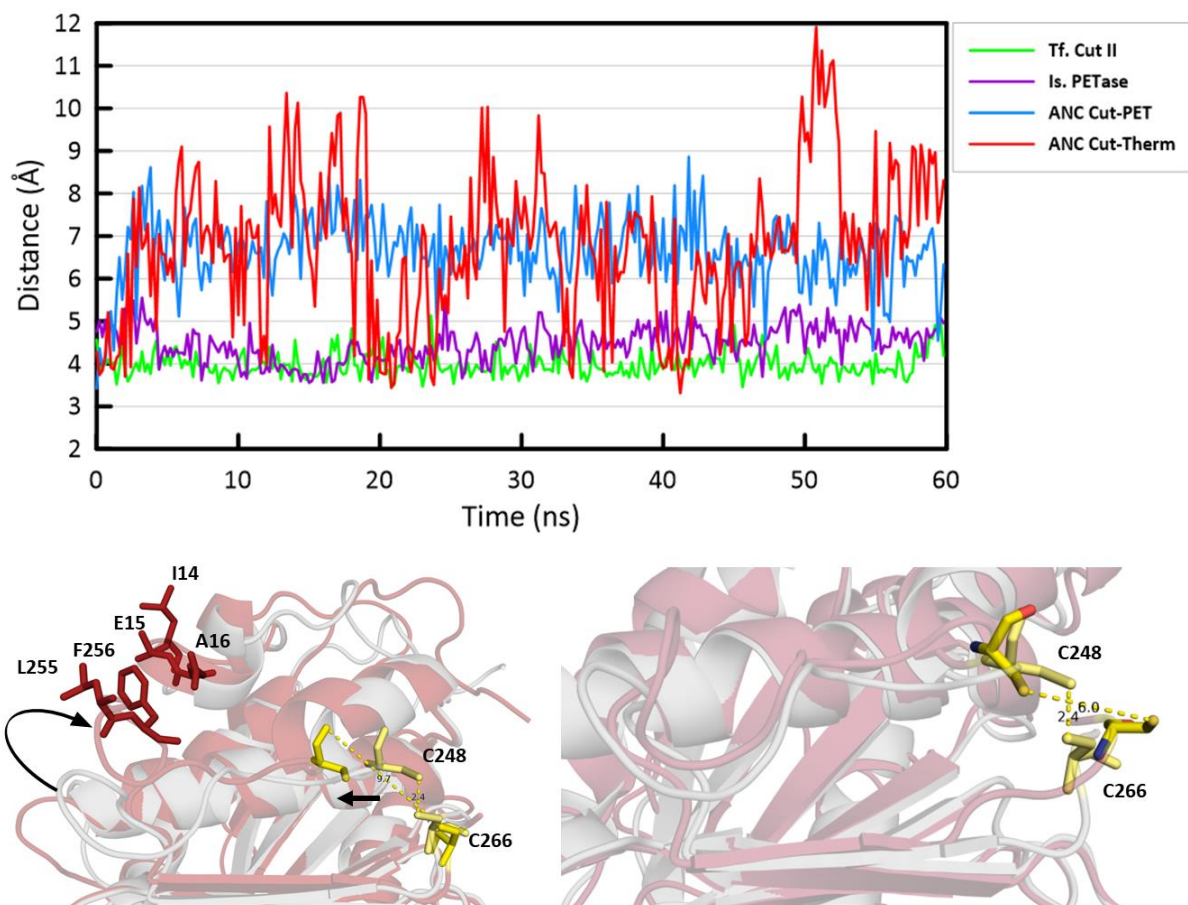


Figure 3.12 Molecular dynamics reveal flexible C-terminal may adversely impact disulfide formation. (Top) Distance of Cys SG atoms during simulation for Tf. Cut II (red), Is. PETase (dark blue), ANC Cut-PETase (green), and ANC Cut-Therm (purple). Distance in Tf. Cut II and Is. PETase remains stable and low throughout the simulation, while both ancestors show increased distances. Further ANC Cut-Therm experiences larger movement scales where loop traverses up towards N-terminal and back throughout the simulations. (Bottom left) Overlay of ANC Cut-Therm simulation frames (white vs. red structure) showing range of motion of CYS loop. Upwards shift of cys loop is met with movement of N-terminal helix toward loop. L255, F256 on the CYS loop and I14, E15, and A16 are shown as red sticks. A hydrophobic interaction between F256 and A16 seems to be transiently formed, as the average distance for this pair in the simulation is between 2.5-5 Å. This interaction could alter placement of C248 and C266, as they move from a 2.4 Å distance to a 9.7 Å distance due to loop rearrangement. (Bottom right) Loop of ANC Cut-PETase is relatively stable, and experiences smaller total movements compared to ANC Cut-Therm. Increased distance of 6.0 from 2.4 Å between C248 and C266 is related to a conformational change rather than large scale movements of the loop backbone.

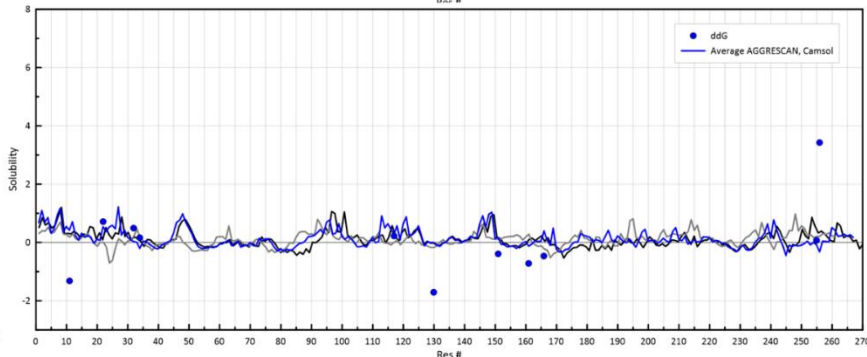
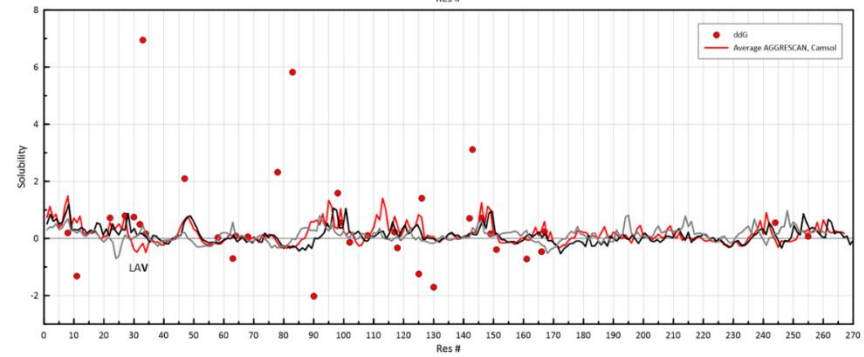
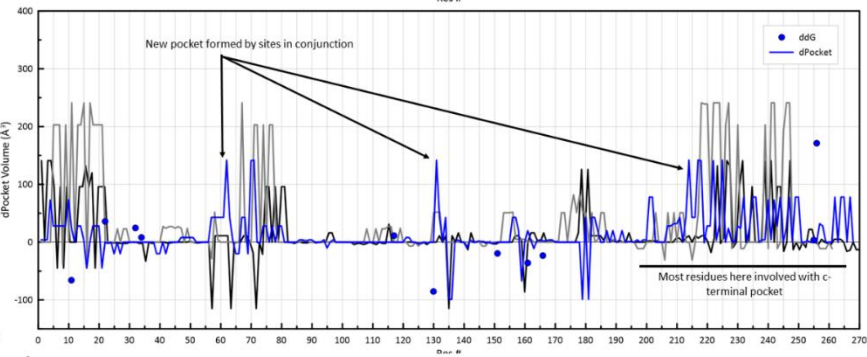
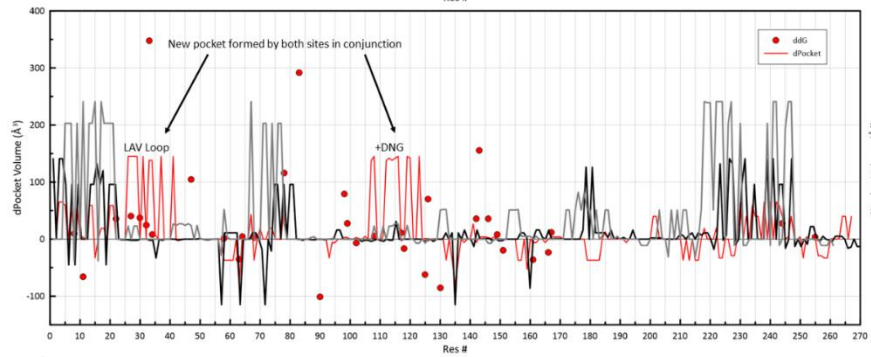
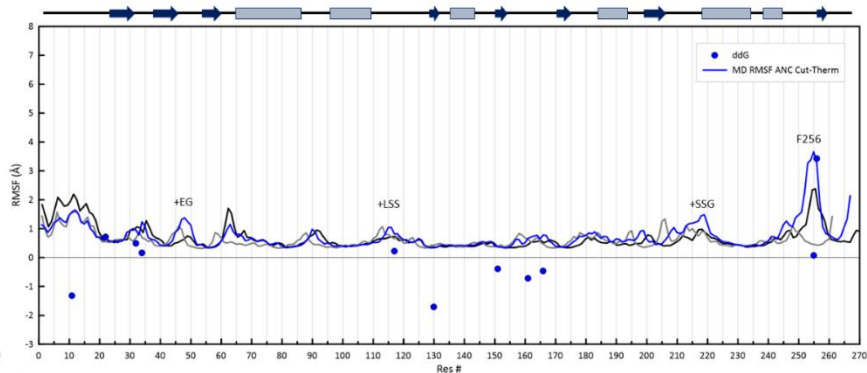
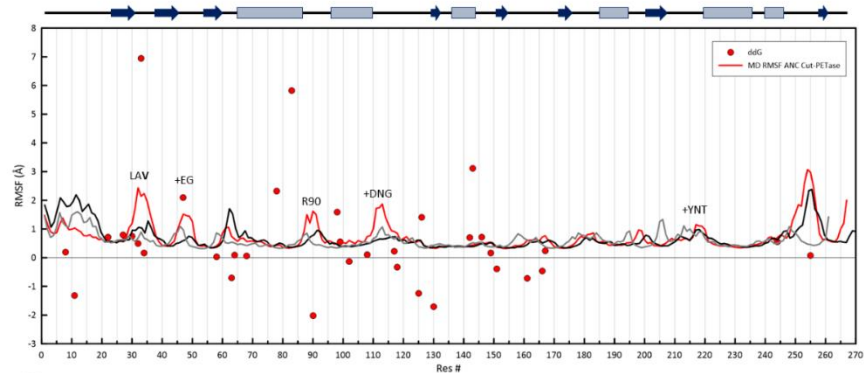
3.8 Synthesis and Summary of Key Sites

The ability to investigate mutational impacts on protein structure can often be challenging, especially when multiple mutations are made simultaneously. Perhaps, then one must *interview* the mutated sequence with multiple approaches, and use the summation of the answers to produce a hypothesis. In the previous sections, bioinformatics-based analysis revealed insert-based changes to loop management that are not represented in ancestors (presence of R120, R122, no gain of INSEQ conserved P118 at Loop #1; no gain of INSEQ conserved disulfide at Loop #2). Further, ddG values of Thh_Est mutation to ancestor residues suggest ANC Cut-Therm mutations are well tolerated, while ANC Cut-PETase may have some significantly destabilizing residues. In experiments it is observed that formation of inter-disulfide linkages is possible on top of intra-disulfide linkages. This was speculated in the previous section to be due to increased flexibility in the C-terminal. Here, all of these observations are combined with C α RMSF (100 ns, average of 3 simulations), residue pocket-involvement (dPocket, which is a measure of changes to protein pocket volume after 100 ns of simulation), and residue solubility in Figure 3.13, in order to understand which areas may be most impactful on ancestors.

Deviation in RMSF to ANC Cut-PETase in relation to Tf. Cut II and Is. PETase are primarily at sites V33, the semi-conserved EG insert, R90, loop insert #1 (DNG), and the C-terminal. Notably, LAV, R90, and loop insert #1 are spatially close, and their movements could affect one another. Motion of LAV and loop insert #1 creates a new, large pocket between the first N-terminal beta sheet and an alpha helix. This opening may be relevant to long-term packing of this area. Furthermore, the LAV loop sequence (which is solvent exposed) is predicted to be aggregation prone by Camsol and AGGRESCAN. As the LAV loop is a site of multiple mutations from Thh_Est and other GAPSEQs, it may be advantageous to reconsider its design. ANC Cut-

PETase seems to have changes on its N-terminal half that are less tolerated than in its C-terminal half. In addition, many changes that had been thought to be considerably destabilizing in Thh_Est, are shown here to have very little impact to their environments in ANC Cut-PETase (for example, A143). However, other changes, such as T47, and V33, which were considered destabilizing in Thh_Est, and correlate to non-wildtype like behaviour here, may truly be destabilizing. Importantly, the active site does not seem to be a starting point for potential problems, though loss of the cHis disulfide should still be considered deactivating in the long-term as per Is. PETase knockouts.

The RMSF, aggregation propensity, and pocket changes in ANC Cut-Therm are very similar to those in Tf. Cut II and Is. PETase. Deviations are prominent at the semi-conserved EG, loop insert #2 (+SSG), and the C-terminal. The movement in cHis seems to cause an opening to the active site, as shown by a new pocket formation that involves several loops in this region. As considered earlier, this flexibility can negatively impact activity, by destabilizing the cHis-cAsp-cSer triad. The opening at the C-terminal for ANC Cut-Therm does not differ in total size from Tf. Cut II, but seems to have more residues involved particularly at the C-terminal loop and nearby beta-sheet, and alpha helix. Unlike, previously shown results for this region, here the disulfide bond is formed, and thus interactions between F256 and A16 are not observed. However, F256 still seems to be a major contributor to flexibility of this region, and is considered a destabilizing mutation relative to Thh_Est. Further other sequences with F256 (Tf. Cut II, Thc_Cut II), all have conserved upstream R-D (S-T in ANC) and a downstream G (S in ANC) which may be necessary to stabilize the loop when F256 is present. Thus, this ancestor reveals problems at the C-terminal, and active sites which may have led to its inactivity and poorer stability.



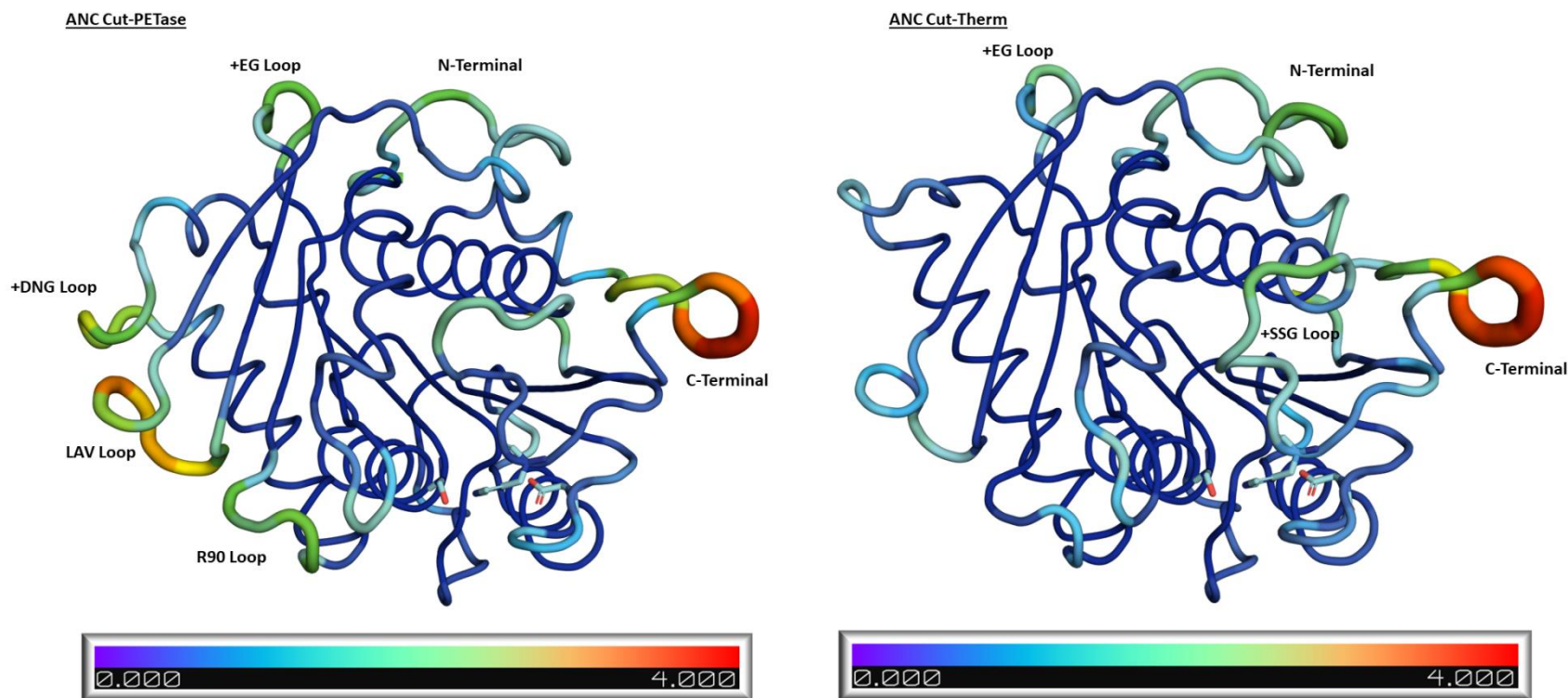


Figure 3.13 Synthesis of results from simulations, pocket analysis, and comparison to Thh Est may aid in determining critical ANC sites. Secondary structure pattern is displayed at the top of the graphs, in a manner similar to previous. For all graphs dots represent point mutations from Thh_Est, and are spread according to their ddG value. Hence, more destabilizing mutations are high, and stabilizing mutations are low. ANC Cut-PETase graphs are all located on the right (shown in red lines/dots), while ANC Cut-Therm are all located on the left (shown in blue lines/dots). In all graphs results for Tf. Cut II is shown in black, while results for Is. PETase is shown in grey. Insert regions are specified by a (+) symbol when labelled. **(Top)** RMSF of ANC compared to Tf. Cut II and Is. PETase at 30 °C. Higher RMSF values correlate to increased fluctuation or dynamics of site. **(Centre)** Changes to pocket volume between 0 ns and 100 ns of simulation. In cases where pockets are created by multiple sites, arrows are used to link them. **(Bottom)** Aggrescan and Camsol average results for ANCs vs. Tf. Cut II and Is. PETase. Here, positive values are associated with solubility, while negative values are associated with increased aggregation propensity. **(Structures)** Secondary structure representations of ANC CUT-PETase (left), and ANC Cut-Therm (right) coloured by average RMSF at 30 °C (3 simulations, 100 ns). This is essentially a visual representation of the RMSF data in the top figure; areas with high RMSF are labelled.

4 Discussion

Reconstructed ancestors often display desirable characteristics, including enhanced thermostability and unique activity profiles (Barruetabeña, Alonso-Lerma, ... Perez-Jimenez, 2019; Gumulya, Baek, ... Gillam, 2018; Whitfield, Zhang, ... Jackson, 2015). The capacity for hyper-stabilization has been in part credited to a “consensus effect” whereby inclusion of common amino acids are preferred over a rare one in an alignment (Porebski and Buckle, 2016; Risso, Gavira, ... Sanchez-Ruiz, 2013; Williams, Pollock, ... Goldstein, 2006). A key distinction, however, is that ASR is capable of considering evolutionary context, and favours residues that are relevant to nearby extant proteins. This is well-demonstrated by ANC Cut-Therm, which has an 80% identity to the alignment consensus, while bolstering a 93% identity to its closest neighbour, Thh_Est. Thereby, choice of ancestral node (depth in tree, activity, and distance of descendants) can modulate the design of sites that are maybe relevant to less conserved properties (such as substrate specificity).

This thesis reports on the design and production of two ancestral alpha-beta hydrolases, specifically with the hope of capturing PET hydrolase activity found in some extant enzymes. In this attempt, loop inserts found in several PET-hydrolases were retained due to their perceived involvement with improving activity against PET (Joo, Cho, ... Kim, 2018). Upon their reconstruction, both ancestors were observed to be less stable, less folded, and inactive against pNPA/pNPP, with ANC Cut-Therm exhibiting slightly better folding and stability results. Assessment of alignment for insert-based perturbations reveal a distinction in loop management between INSEQs and GAPSEQs—specifically by replacement of a helix turn for a proline in loop #1, and addition of a disulfide at loop #2. The inserts represent a major change between ANC Cut-

Therm and Thh_Est, and are hypothesized to have significant involvement in the impaired ANC properties. Indeed, almost all sites considered in the summary (Section 3.8) are loop residues, and are typically shrouded by some uncertainty in the alignment. For example, the EG loop is almost fully present in distant extant sequences from ANCs, while missing in almost all ANC descendants. Loop insert #1, and loop insert #2 are poorly represented in alignments—particularly in ANC Cut-PETase. The LAV loop in Cut-PETase descendants is occasionally missing (R--P, F-A), and may have resulted in a less confident prediction. Finally, the C-terminal loop was inherently gappy, and strange combinations in this area are simply too dynamic (particularly ANC Cut-Therm).

The structure of Tf. Cut II is composed of ~40-50% loops/turns, and their management is a crucial component for the protein's stability and activity. This is especially true, as the catalytic residues and binding site is formed by the intersection of several loops, and a close interaction network between the catalytic triad is required for activity. Thus, its loop structures are unlikely to behave as random coils, and are instead infused with the adequate structure and dynamics for their role. Indeed, many proteins use structured solvent-exposed loops as first line interactors for binding (antibody loops for example, are prestructured for highly specific binding), and for active site protection (i.e. lids) (Zimmermann and Jernigan, 2012). Further, Tf. Cut II contains 14 loops that are responsible for connecting secondary structure elements—some of these are short and simply connect two neighbouring segments, while others are long and transverse the length of the protein. It is likely that their behaviour is important during protein folding as loops and turns can alter the conformational entropy of the unfolded protein, and turn propensity may be an important property for superstructure nucleation (Jappelli, Luzzago, ... Cesareni, 1992; Marcelino and Gierasch, 2008). Thus, while loops are often less predictable and more amenable to site-mutations

in comparison to helices and beta sheets, they play a non-ignorable role in a proteins function, stability, and folding (Guo, Choe, and Loeb, 2004).

In this ASR attempt, modulation of loop length at three sites were ultimately attempted, including two three-residue insertions, and cementation of the C-terminal disulfide loop at eleven residues in length. One of these loops is in a highly sensitive area, and is responsible for housing the catalytic histidine. Thus, any altered conformations of this loop, such that the loop is removed from the active site, can have devastating effects on activity. Likewise, increased flexibility of the C-terminal disulfide loop may be involved with the higher rate of inter-molecular disulfide bond formation seen in ancestors. Numerous investigations have shown that increases to loop length are generally followed by decreased stability possibly due to increased conformation dynamics and entropy of the unfolded state (Dagan, Hagai, ... Reich, 2013; Gavrilov, Dagan, and Levy, 2015; Nagi and Regan, 1997). Thus, one consideration for remediation of ancestral design is to remove loop inserts, and to optimize a shorter C-terminal. However, this may not be ideal, as longer loops, particularly at the active site is still thought to be relevant to function. Further, there may be some unforeseen role of the C-terminal extension that is relevant to stability, as the longer C-terminal is found unproportionally in enzymes from thermostable hosts.

Perhaps, then, given a desire to retain loop inserts, the idea would be to ameliorate loop stabilization. Proteins use three major schemes in order to restrain the behaviour of loop regions: geometric constrains of backbone geometry, interactions with the main protein body, and covalent disulfide linkages. The first is particularly important in regions where structured or rigidified loops are required. Residues with restricted backbone angles (ϕ , ψ), such as proline in combination with sterically accommodating ones, like Gly, can strongly promote and form loop structures (such as turns) (Krieger, Möglich, and Kiefhaber, 2005; Nadeau and Deber, 2016). The gain of a conserved

proline in enzymes with loop-insert #1 in lieu of a short helix turn, may be indicative of this strategy being used to rigidify loop behaviour. Similarly, this may be at play for the more proline rich active site of PET 27 and PET 30, as they support the loop #2 insert without the use of an additional active site disulfide.

Loop interactions with the main body of the protein have also been widely observed in the Tf. Cut II structure. These interactions allow loops to be adhered to a more stable structure, and is especially relevant for long, transversing loops. Additionally, in the terminal hemisphere, arginine residues, that are capable of forming multiple interactions due to their two guanidium groups, are seen forming crosslinks between loops and secondary structure—perhaps gluing distant secondary items together. This function of loops may also help provide a clue into the role of the highly flexible but relatively conserved N terminal, as it is found to contact 4-5 separate loops/strands. In this way, loops may provide a mode of cross-protein communication by network extension between structural gaps.

Disulfide bonds are a more intense form of structural linkage that is employed in the stability of the C-terminal and at loop insert #2 for stabilization (Joo, Cho, ... Kim, 2018; Sulaiman, You, ... Kanaya, 2014). The importance of the disulfide is widely acknowledged, and many engineering attempts on PET hydrolases have involved disulfide addition to infer stability (Oda, Yamagami, ... Kawai, 2018; Then, Wei, ... Zimmermann, 2016; Tournier, Topham, ... Marty, 2020; Zhong-Johnson, Voigt, and Sinskey, 2021). It is speculated that increased thermodynamic stability with disulfide addition is in part due to topological constraints placed on linked regions—thus, many proteins experience some unfolding, and subsequent loss of thermostability without their native disulfides (Klink, Woycechowsky, ... Raines, 2000; Muttathukattil, Singh, and Reddy, 2019). Indeed, it may be possible to assume that malformation

of the C-terminal disulfide in the ancestors leads to reduced stability, and local unfolding. This of course, can lead to aggregation, and colloquially it has been viewed experimentally that reduced Tf. Cut II and ancestors come out of solution more readily. In a similar manner, enzymes with loop insert #2 likely use disulfides to constrain the cHis loop and ensure its proper placement (Fecker, Galaz-Davison, ... Ramírez-Sarmiento, 2018; Joo, Cho, ... Kim, 2018). Thus, disulfides are a promising way of enforcing certain structure. The caveat however, is that large numbers of disulfides can often lead to expression problems, and they result in susceptibility to reducing agents, and scrambling (Berkmen, 2012; Gori, Gagni, and Rinaldi, 2017). Thus, while additions of disulfides are sometimes a necessary requirement for conformational entrapment, the likelihood of rearrangements and potential problems increases with increasing disulfide addition.

Loop engineering in enzymes is primarily completed based on single site mutations, usually in attempts to alter some function such as binding specificity. Ancestral sequence reconstruction is an incredible resource for prediction of loop states related to a subset of a protein family, and has resulted in engineered variants with altered binding properties. One major limitation, however, is that ASR is limited in its ability to accommodate indel based perturbations when a given indel is less common in the ancestral descendants, or in extant proteins distant to a reconstructed ancestor. As the length of the ancestor sequence is set in the first step, perhaps ideally, ASR in these scenarios would be capable of predicting gap existence, which may be particularly helpful for highly gappy regions containing single residue indel events, as is observed in the C-terminal disulfide region of alpha beta hydrolases. In lieu of these capabilities, it is up to researchers to use system specific expertise in order to make sometimes tricky decisions. These limitations are currently at the forefront of ASR algorithm advancement (Foley, Mora, ... Bodén, 2022; Jowkar, Pečerska, ... Anisimova, 2022).

4.1 Future Ideas

Following normal conventions for alignment coverage requirements, the ancestors reconstructed here would not have had loop inserts #1 and #2, which are considered desirable. How then is one to accommodate these? In the alignments, sequence homologs with good sequence similarity, showed insert based perturbations that were highly correlated to the length of the insert, but unrelated to the sequence of the insert itself. Perhaps then, a statistical coupling for insert existence and amino acid identities along the alignment can be determined. In the current design, this strategy would have increased the likelihood of loop accommodation and activity/stability retention by discovery of the insert-induced disulfide bond and proline addition. Further it could reveal anti-correlated perturbations as well, for example removal of R120 in Loop Insert #1. Thus, for this family, it may be ideal to (1) complete ASR first without loop inserts, (2) design loop insert sequences (perhaps based on consensus of loop inserts, or from a desirable extant protein), and (3) rationally adjust sequence for insert based perturbations.

Secondly, it would be ideal to treat ASR with some modularity, especially when the goal is protein engineering rather than elucidation of molecular evolution events. For example, predictions at the relatively variable, gappy C-terminal likely resulted in problems with disulfide formation. Perhaps one method to accommodate this variability then is consensus design in very gappy areas, where only sequences with the desired lengths are considered. This may help prevent aberrant residues from being included, and increase the probability for formation of a critical disulfide bond.

Finally, in section 3.7.2, it was determined that current INSEQs contain 17/25 mutations that have been discovered as activity improving in PET hydrolases. This is in contrast to the 4-5

inherited by the designed ancestors. A neighbour joining tree of known PET hydrolases (Table 1) indeed reveals branching of known PET hydrolases into three groups based on these loop inserts: INSEQs (with and without loop insert #2 disulfide), and GAPSEQs (Figure 4.1). As ASR is extremely related to the properties of the extant descendants, perhaps, reconstruction completed on this smaller subset (INSEQs) may be of particular interest. Notably, a similar splitting was seen by Joo et al., where they further observed that key binding site residues within each group tended to be shared—this additionally led to the suggestion that sequences belonging to the Is. PETase group would naturally have higher activity (Figure 4.1) (Joo, Cho, ... Kim, 2018). However, canonicalization of binding features have not yet been realized, and with the discovery/engineering of novel PET hydrolases in 2021-2022 activity improving residues display more diversity (Chen, Han, ... Guo, 2021). Thus, while the presence of certain residues might make it more likely for improving activity, ancestors generated from these new nodes would need to be probed for PET binding (perhaps by computational docking analysis first, prior to expression). Further, as INSEQ thermostability ranges from 37-60 °C, the consensus effect of ASR may be able to bolster their thermostability while retaining or improving activity. Usage of this smaller subset alone will further relieve the need for post-op loop design for some regions, which may ameliorate possibilities of success. Finally, this thesis has described several computational methods that could be used to assess models prior to experimental expression.

4.2 Conclusions

Ancestral sequence reconstruction is a compelling approach for creating stable and active variants based on existing proteins. This work demonstrates an initial attempt at engineering, which ultimately resulted in less stable, less folded, and inactive variants. Despite significant obstacles in folding and aggregation, both ancestors were successfully expressed and characterized using biophysical methods, which aids considerably in deciphering computational results. The leading hypothesis gained from computational investigation, specifically by comparison of Thh_Est to ANC Cut-Therm (which are 93% identical), is that loop inserts are a likely area of complications. By forcing GAPSEQ like ancestors to accommodate INSEQ like loops, without altering their loop management strategy may have been a critical source of frustration in folding and stability. Further, one of these loops houses the cHIS whose destabilization is detrimental to activity. Thus, extant proteins show a masterly ability to accommodate loop dynamics and structure, and their management strategy changes based on loop length/need to preserve conserved interactions.

Often multi-mutants are created in a semi-rational way, usually relying on bioinformatics to predict tolerance and stability (consensus). Thus, by using nature's trail and error pathway as an example, the opportunities for success increase. That said, protein engineering is an evolving field. This is not the first time, nor will it be the last time that attempts at a multi-mutation enzyme has poor outcomes. This presented work attempts to lay a foundation for strategies to ameliorate ASR outcomes, specifically when gappy regions or indels are present in the alignment.

References

- Alford, R.F., Leaver-Fay, A., Jeliaskov, J.R., O'Meara, M.J., DiMaio, F.P., Park, H., Shapovalov, M.V., Renfrew, P.D., Mulligan, V.K., Kappel, K., Labonte, J.W., Pacella, M.S., Bonneau, R., Bradley, P., Dunbrack, R.L., Das, R., Baker, D., Kuhlman, B., Kortemme, T., Gray, J.J., 2017. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* 13, 3031–3048. <https://doi.org/10.1021/acs.jctc.7b00125>
- Araújo, R., Silva, C., O'Neill, A., Micaelo, N., Guebitz, G., Soares, C.M., Casal, M., Cavaco-Paulo, A., 2007. Tailoring cutinase activity towards polyethylene terephthalate and polyamide 6,6 fibers. *J. Biotechnol.* 128, 849–857. <https://doi.org/10.1016/j.jbiotec.2006.12.028>
- 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., El Omari, 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.* 115. <https://doi.org/10.1073/pnas.1718804115>
- Babkova, P., Sebestova, E., Brezovsky, J., Chaloupkova, R., Damborsky, J., 2017. Ancestral Haloalkane Dehalogenases Show Robustness and Unique Substrate Specificity. *ChemBioChem* 18, 1448–1456. <https://doi.org/10.1002/cbic.201700197>
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., Millán, C., Park, H., Adams, C., Glassman, C.R., DeGiovanni, A., Pereira, J.H., Rodrigues, A.V., van Dijk, A.A., Ebrecht, A.C., Opperman, D.J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M.K., Dalwadi, U., Yip, C.K., Burke, J.E., Garcia, K.C., Grishin, N.V., Adams, P.D., Read, R.J., Baker, D., 2021. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876. <https://doi.org/10.1126/science.abj8754>
- Baker, A.-M.M., Mead, J., 2000. Thermoplastics, in: Harper, C.A. (Ed.), *Modern Plastics Handbook*. McGraw-Hill Education, New York.
- Bar-Rogovsky, H., Huguenmatter, A., Tawfik, D.S., 2013. The Evolutionary Origins of Detoxifying Enzymes. *J. Biol. Chem.* 288, 23914–23927. <https://doi.org/10.1074/jbc.M112.427922>
- Barruetabeña, N., Alonso-Lerma, B., Galera-Prat, A., Joudeh, N., Barandiaran, L., Aldazabal, L., Arbulu, M., Alcalde, M., De Sancho, D., Gavira, J.A., Carrion-Vazquez, M., Perez-Jimenez, R., 2019. Resurrection of efficient Precambrian endoglucanases for lignocellulosic biomass hydrolysis. *Commun. Chem.* 2, 76. <https://doi.org/10.1038/s42004-019-0176-6>
- Berkmen, M., 2012. Production of disulfide-bonded proteins in *Escherichia coli*. *Protein Expr. Purif.* 82, 240–251. <https://doi.org/10.1016/j.pep.2011.10.009>
- Biundo, A., Reich, J., Ribitsch, D., Guebitz, G.M., 2018. Synergistic effect of mutagenesis and truncation to improve a polyesterase from *Clostridium botulinum* for polyester hydrolysis. *Sci. Rep.* 8, 3745. <https://doi.org/10.1038/s41598-018-21825-9>

- Bollinger, A., Thies, S., Knieps-Grünhagen, E., Gertzen, C., Kobus, S., Höppner, A., Ferrer, M., Gohlke, H., Smits, S.H.J., Jaeger, K.-E., 2020. A Novel Polyester Hydrolase From the Marine Bacterium *Pseudomonas aestusnigri* – Structural and Functional Insights. *Front. Microbiol.* 11, 114. <https://doi.org/10.3389/fmicb.2020.00114>
- Bonneau, R., Tsai, J., Ruczinski, I., Chivian, D., Rohl, C., Strauss, C.E.M., Baker, D., 2001. Rosetta in CASP4: Progress in ab initio protein structure prediction. *Proteins Struct. Funct. Genet.* 45, 119–126. <https://doi.org/10.1002/prot.1170>
- Brahney, J., Hallerud, M., Heim, E., Hahnenberger, M., 2020. Title: Plastic Rain in Protected Areas of the United States. *Science* 368, 13. <https://doi.org/10.1126/science.aaz5819>
- Broom, A., 2010. From Peptides to Proteins: Exploring Modular Evolution Through the β -Trefoil Fold (Thesis). University of Waterloo.
- Broom, R.A., 2016. Computational Design of Protein Structure and Prediction of Ligand Binding (Thesis). University of Waterloo.
- Brott, S., Pfaff, L., Schuricht, J., Schwarz, J., Böttcher, D., Badenhorst, C.P.S., Wei, R., Bornscheuer, U.T., 2022. Engineering and evaluation of thermostable *Is* PETase variants for PET degradation. *Eng. Life Sci.* 22, 192–203. <https://doi.org/10.1002/elsc.202100105>
- Buchner, J., Kiefhaber, T., 2005. Protein folding handbook. Wiley-VCH, Weinheim.
- Case, A., Belfon, L., Ben-Shalom, S.R., Brozell, D.S., Cerutti, T.E., Cheatham, I.V.W.D., 2020. Amber 2020.
- Case, D.A., H. Metin Aktulga, Kellon Belfon, Ben-Shalom, I.Y., Berryman, J.T., Brozell, S.R., Cerutti, D.S., Cheatham, T.E., G. Andrés Cisneros, Vinícius Wilian D Cruzeiro, Darden, T.A., Duke, R.E., Giambasu, G., Gilson, M.K., Gohlke, H., Goetz, A.W., Harris, R., Izadi, S., Izmailov, S.A., Koushik Kasavajhala, Kaymak, M.C., King, E., Kovalenko, A., Kurtzman, T., Taisung Lee, LeGrand, S., Pengfei Li, Lin, C., Liu, J., Luchko, T., Luo, R., Machado, M., Man, V., Madushanka Manathunga, Merz, K.M., Yinglong Miao, Mikhailovskii, O., Monard, G., Nguyen, H., O’Hearn, K.A., Onufriev, A., Pan, F., Pantano, S., Ruxi Qi, Rahnamoun, A., Roe, D.R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Akhil Shajan, Shen, J., Simmerling, C.L., Skrynnikov, N.R., Smith, J., Swails, J., Walker, R.C., Wang, J., Junmei Wang, Haixin Wei, Wolf, R.M., Xiongwu Wu, Yeyue Xiong, Xue, Y., York, D.M., Shiji Zhao, Kollman, P.A., 2022. Amber 2022. <https://doi.org/10.13140/RG.2.2.31337.77924>
- Chamas, A., Moon, H., Zheng, J., Qiu, Y., Tabassum, T., Jang, J.H., Abu-Omar, M., Scott, S.L., Suh, S., 2020. Degradation Rates of Plastics in the Environment. *ACS Sustain. Chem. Eng.* 8, 3494–3511. <https://doi.org/10.1021/acssuschemeng.9b06635>
- Charupanit, K., Tipmanee, V., Sutthibutpong, T., Limsakul, P., 2022. In Silico Identification of Potential Sites for a Plastic-Degrading Enzyme by a Reverse Screening through the Protein Sequence Space and Molecular Dynamics Simulations. *Molecules* 27, 3353. <https://doi.org/10.3390/molecules27103353>
- Chen, C.-C., Han, X., Li, X., Jiang, P., Niu, D., Ma, L., Liu, W., Li, S., Qu, Y., Hu, H., Min, J., Yang, Y., Zhang, L., Zeng, W., Huang, J.-W., Dai, L., Guo, R.-T., 2021. General features to enhance enzymatic activity of poly(ethylene terephthalate) hydrolysis. *Nat. Catal.* 4, 425–430. <https://doi.org/10.1038/s41929-021-00616-y>
- Chen, K., Hu, Y., Dong, X., Sun, Y., 2021. Molecular Insights into the Enhanced Performance of EKylated PETase Toward PET Degradation. *ACS Catal.* 11, 7358–7370. <https://doi.org/10.1021/acscatal.1c01062>

- Chen, K., Quan, M., Dong, X., Shi, Q., Sun, Y., 2021. Low modification of PETase enhances its activity toward degrading PET: Effect of conjugate monomer property. *Biochem. Eng. J.* 175, 108151. <https://doi.org/10.1016/j.bej.2021.108151>
- Clifton, B.E., Kaczmarek, J.A., Carr, P.D., Gerth, M.L., Tokuriki, N., Jackson, C.J., 2018. Evolution of cyclohexadienyl dehydratase from an ancestral solute-binding protein. *Nat. Chem. Biol.* 14, 542–547. <https://doi.org/10.1038/s41589-018-0043-2>
- Conchillo-Solé, O., de Groot, N.S., Avilés, F.X., Vendrell, J., Daura, X., Ventura, S., 2007. AGGRESCAN: a server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. *BMC Bioinformatics* 8, 65. <https://doi.org/10.1186/1471-2105-8-65>
- Cui, Y., Chen, Y., Liu, X., Dong, S., Tian, Y., Qiao, Y., Mitra, R., Han, J., Li, C., Han, X., Liu, W., Chen, Q., Wei, W., Wang, X., Du, W., Tang, S., Xiang, H., Liu, H., Liang, Y., Houk, K.N., Wu, B., 2021. Computational Redesign of a PETase for Plastic Biodegradation under Ambient Condition by the GRAPE Strategy. *ACS Catal.* 11, 1340–1350. <https://doi.org/10.1021/acscatal.0c05126>
- da Silva, D.J., Wiebeck, H., 2020. Current options for characterizing, sorting, and recycling polymeric waste. *Prog. Rubber Plast. Recycl. Technol.* 36, 284–303. <https://doi.org/10.1177/1477760620918603>
- Dagan, S., Hagai, T., Gavrilov, Y., Kapon, R., Levy, Y., Reich, Z., 2013. Stabilization of a protein conferred by an increase in folded state entropy. *Proc. Natl. Acad. Sci.* 110, 10628–10633. <https://doi.org/10.1073/pnas.1302284110>
- Danso, D., Schmeisser, C., Chow, J., Zimmermann, W., Wei, R., Leggewie, C., Li, X., Hazen, T., Streit, W.R., 2018. New Insights into the Function and Global Distribution of Polyethylene Terephthalate (PET)-Degrading Bacteria and Enzymes in Marine and Terrestrial Metagenomes. *Appl. Environ. Microbiol.* 84, e02773-17. <https://doi.org/10.1128/AEM.02773-17>
- Dauvergne, P., 2018. Why is the global governance of plastic failing the oceans? *Glob. Environ. Change* 51, 22–31. <https://doi.org/10.1016/j.gloenvcha.2018.05.002>
- Deng, Y., Zhang, Y., Lemos, B., Ren, H., 2017. Tissue accumulation of microplastics in mice and biomarker responses suggest widespread health risks of exposure. *Sci. Rep.* 7, 46687. <https://doi.org/10.1038/srep46687>
- Derraik, J.G.B., 2002. The pollution of the marine environment by plastic debris: a review. *Mar. Pollut. Bull.* 44, 842–852. [https://doi.org/10.1016/S0025-326X\(02\)00220-5](https://doi.org/10.1016/S0025-326X(02)00220-5)
- Fecker, T., Galaz-Davison, P., Engelberger, F., Narui, Y., Sotomayor, M., Parra, L.P., Ramírez-Sarmiento, C.A., 2018. Active Site Flexibility as a Hallmark for Efficient PET Degradation by *I. sakaiensis* PETase. *Biophys. J.* 114, 1302–1312. <https://doi.org/10.1016/j.bpj.2018.02.005>
- Foley, G., Mora, A., Ross, C.M., Bottoms, S., Sützl, L., Lamprecht, M.L., Zaugg, J., Essebier, A., Balderson, B., Newell, R., Thomson, R.E.S., Kobe, B., Barnard, R.T., Guddat, L., Schenk, G., Carsten, J., Gumulya, Y., Rost, B., Haltrich, D., Sieber, V., Gillam, E.M.J., Bodén, M., 2022. Engineering indel and substitution variants of diverse and ancient enzymes using Graphical Representation of Ancestral Sequence Predictions (GRASP). *PLOS Comput. Biol.* 18, e1010633. <https://doi.org/10.1371/journal.pcbi.1010633>
- Furukawa, M., Kawakami, N., Oda, K., Miyamoto, K., 2018. Acceleration of Enzymatic Degradation of Poly(ethylene terephthalate) by Surface Coating with Anionic Surfactants. *ChemSusChem* 11, 4018–4025. <https://doi.org/10.1002/cssc.201802096>

- Furukawa, M., Kawakami, N., Tomizawa, A., Miyamoto, K., 2019. Efficient Degradation of Poly(ethylene terephthalate) with *Thermobifida fusca* Cutinase Exhibiting Improved Catalytic Activity Generated using Mutagenesis and Additive-based Approaches. *Sci. Rep.* 9, 16038. <https://doi.org/10.1038/s41598-019-52379-z>
- Gamerith, C., Vastano, M., Ghorbanpour, S.M., Zitzenbacher, S., Ribitsch, D., Zumstein, M.T., Sander, M., Herrero Acero, E., Pellis, A., Guebitz, G.M., 2017. Enzymatic Degradation of Aromatic and Aliphatic Polyesters by *P. pastoris* Expressed Cutinase 1 from *Thermobifida cellulositytica*. *Front. Microbiol.* 8, 938. <https://doi.org/10.3389/fmicb.2017.00938>
- Gamiz-Arco, G., Gutierrez-Rus, L.I., Risso, V.A., Ibarra-Molero, B., Hoshino, Y., Petrović, D., Justicia, J., Cuerva, J.M., Romero-Rivera, A., Seelig, B., Gavira, J.A., Kamerlin, S.C.L., Gaucher, E.A., Sanchez-Ruiz, J.M., 2021. Heme-binding enables allosteric modulation in an ancient TIM-barrel glycosidase. *Nat. Commun.* 12, 380. <https://doi.org/10.1038/s41467-020-20630-1>
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server, in: Walker, J.M. (Ed.), *The Proteomics Protocols Handbook*. Humana Press, Totowa, NJ, pp. 571–607. <https://doi.org/10.1385/1-59259-890-0:571>
- Gavrilov, Y., Dagan, S., Levy, Y., 2015. Shortening a loop can increase protein native state entropy: Modulating Protein Native State Entropy. *Proteins Struct. Funct. Bioinforma.* 83, 2137–2146. <https://doi.org/10.1002/prot.24926>
- Gercke, D., Furtmann, C., Tozakidis, I.E.P., Jose, J., 2021. Highly Crystalline Post-Consumer PET Waste Hydrolysis by Surface Displayed PETase Using a Bacterial Whole-Cell Biocatalyst. *ChemCatChem* 13, 3479–3489. <https://doi.org/10.1002/cctc.202100443>
- Geyer, R., Jambeck, J.R., Law, K.L., 2017. Production, use, and fate of all plastics ever made. *Sci. Adv.* 3, e1700782. <https://doi.org/10.1126/sciadv.1700782>
- Ghatge, S., Yang, Y., Ahn, J.-H., Hur, H.-G., 2020. Biodegradation of polyethylene: a brief review. *Appl. Biol. Chem.* 63, 27. <https://doi.org/10.1186/s13765-020-00511-3>
- Goldenzweig, A., Goldsmith, M., Hill, S.E., Gertman, O., Laurino, P., Ashani, Y., Dym, O., Unger, T., Albeck, S., Prilusky, J., Lieberman, R.L., Aharoni, A., Silman, I., Sussman, J.L., Tawfik, D.S., Fleishman, S.J., 2016. Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Mol. Cell* 63, 337–346. <https://doi.org/10.1016/j.molcel.2016.06.012>
- Gori, A., Gagni, P., Rinaldi, S., 2017. Disulfide Bond Mimetics: Strategies and Challenges. *Chem. - Eur. J.* 23, 14987–14995. <https://doi.org/10.1002/chem.201703199>
- Gromiha, M.M., Selvaraj, S., 2001. Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction11Edited by P. E. Wright. *J. Mol. Biol.* 310, 27–32. <https://doi.org/10.1006/jmbi.2001.4775>
- Groussin, M., Daubin, V., Gouy, M., Tannier, E., 2016. Ancestral Reconstruction: Theory and Practice, in: *Encyclopedia of Evolutionary Biology*. Elsevier, pp. 70–77. <https://doi.org/10.1016/B978-0-12-800049-6.00166-9>
- Guan, Y., Zhu, Q., Huang, D., Zhao, S., Jan Lo, L., Peng, J., 2015. An equation to estimate the difference between theoretically predicted and SDS PAGE-displayed molecular weights for an acidic peptide. *Sci. Rep.* 5, 13370. <https://doi.org/10.1038/srep13370>

- Gumulya, Y., Baek, J.-M., Wun, S.-J., Thomson, R.E.S., Harris, K.L., Hunter, D.J.B., Behrendorff, J.B.Y.H., Kulig, J., Zheng, S., Wu, X., Wu, B., Stok, J.E., De Voss, J.J., Schenk, G., Jurva, U., Andersson, S., Isin, E.M., Bodén, M., Guddat, L., Gillam, E.M.J., 2018. Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nat. Catal.* 1, 878–888. <https://doi.org/10.1038/s41929-018-0159-5>
- Guo, B., Vanga, S.R., Lopez-Lorenzo, X., Saenz-Mendez, P., Ericsson, S.R., Fang, Y., Ye, X., Schriever, K., Bäckström, E., Biundo, A., Zubarev, R.A., Furó, I., Hakkarainen, M., Syrén, P.-O., 2022. Conformational Selection in Biocatalytic Plastic Degradation by PETase. *ACS Catal.* 12, 3397–3409. <https://doi.org/10.1021/acscatal.1c05548>
- Guo, H.H., Choe, J., Loeb, L.A., 2004. Protein tolerance to random amino acid change. *Proc. Natl. Acad. Sci.* 101, 9205–9210. <https://doi.org/10.1073/pnas.0403255101>
- Guruprasad, K., Reddy, B.V.B., Pandit, M.W., 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Protein Eng. Des. Sel.* 4, 155–161. <https://doi.org/10.1093/protein/4.2.155>
- Han, X., Liu, W., Huang, J.-W., Ma, J., Zheng, Y., Ko, T.-P., Xu, L., Cheng, Y.-S., Chen, C.-C., Guo, R.-T., 2017. Structural insight into catalytic mechanism of PET hydrolase. *Nat. Commun.* 8, 2106. <https://doi.org/10.1038/s41467-017-02255-z>
- Hegde, K., Dasu, V.V., 2014. Structural Stability and Unfolding Properties of Cutinases from *Thermobifida fusca*. *Appl. Biochem. Biotechnol.* 174, 803–819. <https://doi.org/10.1007/s12010-014-1037-5>
- Hendrikse, N.M., Charpentier, G., Nordling, E., Syrén, P., 2018. Ancestral diterpene cyclases show increased thermostability and substrate acceptance. *FEBS J.* 285, 4660–4673. <https://doi.org/10.1111/febs.14686>
- Herrero Acero, E., Ribitsch, D., Dellacher, A., Zitzenbacher, S., Marold, A., Steinkellner, G., Gruber, K., Schwab, H., Guebitz, G.M., 2013. Surface engineering of a cutinase from *Thermobifida cellulositytica* for improved polyester hydrolysis: Surface Engineering of a Cutinase. *Biotechnol. Bioeng.* 110, 2581–2590. <https://doi.org/10.1002/bit.24930>
- Herrero Acero, E., Ribitsch, D., Steinkellner, G., Gruber, K., Greimel, K., Eiteljoerg, I., Trotscha, E., Wei, R., Zimmermann, W., Zinn, M., Cavaco-Paulo, A., Freddi, G., Schwab, H., Guebitz, G., 2011. Enzymatic Surface Hydrolysis of PET: Effect of Structural Diversity on Kinetic Properties of Cutinases from *Thermobifida*. *Macromolecules* 44, 4632–4640. <https://doi.org/10.1021/ma200949p>
- Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35, 518–522. <https://doi.org/10.1093/molbev/msx281>
- Hosaka, M., Kamimura, N., Toribami, S., Mori, K., Kasai, D., Fukuda, M., Masai, E., 2013. Novel Tripartite Aromatic Acid Transporter Essential for Terephthalate Uptake in *Comamonas* sp. Strain E6. *Appl. Environ. Microbiol.* 79, 6148–6155. <https://doi.org/10.1128/AEM.01600-13>
- Hu, X., Thumarat, U., Zhang, X., Tang, M., Kawai, F., 2010. Diversity of polyester-degrading bacteria in compost and molecular analysis of a thermoactive esterase from *Thermobifida alba* AHK119. *Appl. Microbiol. Biotechnol.* 87, 771–779. <https://doi.org/10.1007/s00253-010-2555-x>
- Huang, X., Cao, L., Qin, Z., Li, S., Kong, W., Liu, Y., 2018. Tat-Independent Secretion of Polyethylene Terephthalate Hydrolase PETase in *Bacillus subtilis* 168 Mediated by Its

- Native Signal Peptide. *J. Agric. Food Chem.* 66, 13217–13227.
<https://doi.org/10.1021/acs.jafc.8b05038>
- Hultqvist, G., Åberg, E., Camilloni, C., Sundell, G.N., Andersson, E., Dogan, J., Chi, C.N., Vendruscolo, M., Jemth, P., 2017. Emergence and evolution of an interaction between intrinsically disordered proteins. *eLife* 6, e16059. <https://doi.org/10.7554/eLife.16059>
- Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: Visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
- Idumah, C.I., 2022. Recent advancements in thermolysis of plastic solid wastes to liquid fuel. *J. Therm. Anal. Calorim.* 147, 3495–3508. <https://doi.org/10.1007/s10973-021-10776-5>
- Ikai, A., 1980. Thermostability and aliphatic index of globular proteins. *J. Biochem. (Tokyo)* 88, 1895–1898.
- Ivankov, D.N., Garbuzynskiy, S.O., Alm, E., Plaxco, K.W., Baker, D., Finkelstein, A.V., 2003. Contact order revisited: Influence of protein size on the folding rate. *Protein Sci.* 12, 2057–2062. <https://doi.org/10.1110/ps.0302503>
- Jappelli, R., Luzzago, A., Tataseo, P., Pernice, I., Cesareni, G., 1992. Loop Mutations can cause a substantial conformational change in the carboxy terminus of the ferritin protein. *J. Mol. Biol.* 227, 532–543. [https://doi.org/10.1016/0022-2836\(92\)90905-Y](https://doi.org/10.1016/0022-2836(92)90905-Y)
- Jia, Y., Samak, N.A., Hao, X., Chen, Z., Yang, G., Zhao, X., Mu, T., Yang, M., Xing, J., 2021. Nano-immobilization of PETase enzyme for enhanced polyethylene terephthalate biodegradation. *Biochem. Eng. J.* 176, 108205. <https://doi.org/10.1016/j.bej.2021.108205>
- Jo, B.H., 2021. An Intrinsically Disordered Peptide Tag that Confers an Unusual Solubility to Aggregation-Prone Proteins. *Appl. Ind. Microbiol.* 88. <https://doi.org/10.1128/aem.00097-22>
- Joo, S., Cho, I.J., Seo, H., Son, H.F., Sagong, H.-Y., Shin, T.J., Choi, S.Y., Lee, S.Y., Kim, K.-J., 2018. Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation. *Nat. Commun.* 9, 382. <https://doi.org/10.1038/s41467-018-02881-1>
- Jowkar, G., Pečerska, J., Maiolo, M., Gil, M., Anisimova, M., 2022. ARPIP: Ancestral Sequence Reconstruction with Insertions and Deletions under the Poisson Indel Process. *Syst. Biol.* syac050. <https://doi.org/10.1093/sysbio/syac050>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohli, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermini, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. <https://doi.org/10.1038/nmeth.4285>
- Katoh, K., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. <https://doi.org/10.1093/nar/gkf436>
- Kawabata, T., Oda, M., Kawai, F., 2017. Mutational analysis of cutinase-like enzyme, Cut190, based on the 3D docking structure with model compounds of polyethylene terephthalate. *J. Biosci. Bioeng.* 124, 28–35. <https://doi.org/10.1016/j.jbiosc.2017.02.007>

- Kawai, F., Kawabata, T., Oda, M., 2019. Current knowledge on enzymatic PET degradation and its possible application to waste stream management and other fields. *Appl. Microbiol. Biotechnol.* 103, 4253–4268. <https://doi.org/10.1007/s00253-019-09717-y>
- Kawai, F., Oda, M., Tamashiro, T., Waku, T., Tanaka, N., Yamamoto, M., Mizushima, H., Miyakawa, T., Tanokura, M., 2014. A novel Ca²⁺-activated, thermostabilized polyesterase capable of hydrolyzing polyethylene terephthalate from *Saccharomonospora viridis* AHK190. *Appl. Microbiol. Biotechnol.* 98, 10053–10064. <https://doi.org/10.1007/s00253-014-5860-y>
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. <https://doi.org/10.1038/nprot.2015.053>
- Klink, T.A., Woycechowsky, K.J., Taylor, K.M., Raines, R.T., 2000. Contribution of disulfide bonds to the conformational stability and catalytic activity of ribonuclease A: Disulfide bonds of ribonuclease A. *Eur. J. Biochem.* 267, 566–572. <https://doi.org/10.1046/j.1432-1327.2000.01037.x>
- Knott, B.C., Erickson, E., Allen, M.D., Gado, J.E., Graham, R., Kearns, F.L., Pardo, I., Topuzlu, E., Anderson, J.J., Austin, H.P., Dominick, G., Johnson, C.W., Rorrer, N.A., Szostkiewicz, C.J., Copié, V., Payne, C.M., Woodcock, H.L., Donohoe, B.S., Beckham, G.T., McGeehan, J.E., 2020. Characterization and engineering of a two-enzyme system for plastics depolymerization. *Proc. Natl. Acad. Sci.* 117, 25476–25485. <https://doi.org/10.1073/pnas.2006753117>
- Ko, H., Kang, M., Kim, M.-J., Yi, J., Kang, J., Bae, J.-H., Sohn, J.-H., Sung, B.H., 2021. A novel protein fusion partner, carbohydrate-binding module family 66, to enhance heterologous protein expression in *Escherichia coli*. *Microb. Cell Factories* 20, 232. <https://doi.org/10.1186/s12934-021-01725-w>
- Kozłowski, L.P., 2016. IPC – Isoelectric Point Calculator. *Biol. Direct* 11, 55. <https://doi.org/10.1186/s13062-016-0159-9>
- Krieger, F., Möglich, A., Kiefhaber, T., 2005. Effect of Proline and Glycine Residues on Dynamics and Barriers of Loop Formation in Polypeptide Chains. *J. Am. Chem. Soc.* 127, 3346–3352. <https://doi.org/10.1021/ja042798i>
- Laist, D.W., 1997. Impacts of Marine Debris: Entanglement of Marine Life in Marine Debris Including a Comprehensive List of Species with Entanglement and Ingestion Records, in: Coe, J.M., Rogers, D.B. (Eds.), *Marine Debris*, Springer Series on Environmental Management. Springer New York, New York, NY, pp. 99–139. https://doi.org/10.1007/978-1-4613-8486-1_10
- Le, S.Q., Gascuel, O., 2008. An Improved General Amino Acid Replacement Matrix. *Mol. Biol. Evol.* 25, 1307–1320. <https://doi.org/10.1093/molbev/msn067>
- Leal Filho, W., Saari, U., Fedoruk, M., Iital, A., Moora, H., Klöga, M., Voronova, V., 2019. An overview of the problems posed by plastic products and the role of extended producer responsibility in Europe. *J. Clean. Prod.* 214, 550–558. <https://doi.org/10.1016/j.jclepro.2018.12.256>
- Lear, G., Maday, S.D.M., Gambarini, V., Northcott, G., Abbel, R., Kingsbury, J.M., Weaver, L., Wallbank, J.A., Pantos, O., 2022. Microbial abilities to degrade global environmental plastic polymer waste are overstated. *Environ. Res. Lett.* 17, 043002. <https://doi.org/10.1088/1748-9326/ac59a7>

- Lebreton, L., Andrady, A., 2019. Future scenarios of global plastic waste generation and disposal. *Palgrave Commun.* 5, 6. <https://doi.org/10.1057/s41599-018-0212-7>
- Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P., Chatonnet, A., 2012. ESTHER, the database of the α/β -hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Res.* 41, D423–D429. <https://doi.org/10.1093/nar/gks1154>
- Leslie, H.A., van Velzen, M.J.M., Brandsma, S.H., Vethaak, A.D., Garcia-Vallejo, J.J., Lamoree, M.H., 2022. Discovery and quantification of plastic particle pollution in human blood. *Environ. Int.* 163, 107199. <https://doi.org/10.1016/j.envint.2022.107199>
- Lin, Z., Jin, T., Zou, T., Xu, L., Xi, B., Xu, D., He, J., Xiong, L., Tang, C., Peng, J., Zhou, Y., Fei, J., 2022. Current progress on plastic/microplastic degradation: Fact influences and mechanism. *Environ. Pollut.* 304, 119159. <https://doi.org/10.1016/j.envpol.2022.119159>
- Liu, B., He, L., Wang, L., Li, T., Li, C., Liu, H., Luo, Y., Bao, R., 2018. Protein Crystallography and Site-Direct Mutagenesis Analysis of the Poly(ethylene terephthalate) Hydrolase PETase from *Ideonella sakaiensis*. *ChemBioChem* 19, 1471–1475. <https://doi.org/10.1002/cbic.201800097>
- Liu, Y., Liu, Z., Guo, Z., Yan, T., Jin, C., Wu, J., 2022. Enhancement of the degradation capacity of IsPETase for PET plastic degradation by protein engineering. *Sci. Total Environ.* 834, 154947. <https://doi.org/10.1016/j.scitotenv.2022.154947>
- Lu, H., Diaz, D.J., Czarnecki, N.J., Zhu, C., Kim, W., Shroff, R., Acosta, D.J., Alexander, B., Cole, H., Zhang, Y.J., Lynd, N., Ellington, A.D., Alper, H.S., 2022. Deep learning redesign of PETase for practical PET degrading applications. *Nature* 604, 662–667. <https://doi.org/10.1038/s41586-022-04599-z>
- Ma, Y., Yao, M., Li, B., Ding, M., He, B., Chen, S., Zhou, X., Yuan, Y., 2018. Enhanced Poly(ethylene terephthalate) Hydrolase Activity by Protein Engineering. *Engineering* 4, 888–893. <https://doi.org/10.1016/j.eng.2018.09.007>
- Madeira, F., Pearce, M., Tivey, A.R.N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., Lopez, R., 2022. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* 50, W276–W279. <https://doi.org/10.1093/nar/gkac240>
- Mädler, S., Barylyuk, K., Boeri Erba, E., Nieckarz, R.J., Zenobi, R., 2012. Compelling Advantages of Negative Ion Mode Detection in High-Mass MALDI-MS for Homomeric Protein Complexes. *J. Am. Soc. Mass Spectrom.* 23, 213–224. <https://doi.org/10.1007/s13361-011-0274-x>
- Maisels, A., Hiller, A., Simon, F., 2022. Chemical Recycling for Plastic Waste: Status and Perspectives. *ChemBioEng Rev.* 9, 541–555. <https://doi.org/10.1002/cben.202200024>
- Marcelino, A.M.C., Gierasch, L.M., 2008. Roles of β -turns in protein folding: From peptide models to protein engineering. *Biopolymers* 89, 380–391. <https://doi.org/10.1002/bip.20960>
- Marchot, P., Chatonnet, A., 2012. Enzymatic Activity and Protein Interactions in Alpha/Beta Hydrolase Fold Proteins: Moonlighting Versus Promiscuity. *Protein Pept. Lett.* 19, 132–143. <https://doi.org/10.2174/092986612799080284>
- Meng, X., Yang, L., Liu, H., Li, Q., Xu, G., Zhang, Yan, Guan, F., Zhang, Yuhong, Zhang, W., Wu, N., Tian, J., 2021. Protein engineering of stable IsPETase for PET plastic degradation by Premuse. *Int. J. Biol. Macromol.* 180, 667–676. <https://doi.org/10.1016/j.ijbiomac.2021.03.058>

- Meyer Cifuentes, I.E., Wu, P., Zhao, Y., Liu, W., Neumann-Schaal, M., Pfaff, L., Barys, J., Li, Z., Gao, J., Han, X., Bornscheuer, U.T., Wei, R., Öztürk, B., 2022. Molecular and Biochemical Differences of the Tandem and Cold-Adapted PET Hydrolases Ple628 and Ple629, Isolated From a Marine Microbial Consortium. *Front. Bioeng. Biotechnol.* 10, 930140. <https://doi.org/10.3389/fbioe.2022.930140>
- Micsonai, A., Moussong, É., Wien, F., Boros, E., Vadász, H., Murvai, N., Lee, Y.-H., Molnár, T., Réfrégiers, M., Goto, Y., Tantos, Á., Kardos, J., 2022. BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy. *Nucleic Acids Res.* 50, W90–W98. <https://doi.org/10.1093/nar/gkac345>
- Miyakawa, T., Mizushima, H., Ohtsuka, J., Oda, M., Kawai, F., Tanokura, M., 2015. Structural basis for the Ca²⁺-enhanced thermostability and activity of PET-degrading cutinase-like enzyme from *Saccharomonospora viridis* AHK190. *Appl. Microbiol. Biotechnol.* 99, 4297–4307. <https://doi.org/10.1007/s00253-014-6272-8>
- Mohanani, N., Montazer, Z., Sharma, P.K., Levin, D.B., 2020. Microbial and Enzymatic Degradation of Synthetic Plastics. *Front. Microbiol.* 11, 580709. <https://doi.org/10.3389/fmicb.2020.580709>
- Monsellier, E., Bedouelle, H., 2005. Quantitative measurement of protein stability from unfolding equilibria monitored with the fluorescence maximum wavelength. *Protein Eng. Des. Sel.* 18, 445–456. <https://doi.org/10.1093/protein/gzi046>
- Muttathukattil, A.N., Singh, P.C., Reddy, G., 2019. Role of Disulfide Bonds and Topological Frustration in the Kinetic Partitioning of Lysozyme Folding Pathways. *J. Phys. Chem. B* 123, 3232–3241. <https://doi.org/10.1021/acs.jpcc.9b00739>
- Nadeau, V.G., Deber, C.M., 2016. Structural impact of proline mutations in the loop region of an ancestral membrane protein: Structural Impact of Proline Mutations. *Biopolymers* 106, 37–42. <https://doi.org/10.1002/bip.22765>
- Nagi, A.D., Regan, L., 1997. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold. Des.* 2, 67–75. [https://doi.org/10.1016/S1359-0278\(97\)00007-2](https://doi.org/10.1016/S1359-0278(97)00007-2)
- Nakamura, A., Kobayashi, N., Koga, N., Iino, R., 2021. Positive Charge Introduction on the Surface of Thermostabilized PET Hydrolase Facilitates PET Binding and Degradation. *ACS Catal.* 11, 8550–8564. <https://doi.org/10.1021/acscatal.1c01204>
- Nakano, S., Minamino, Y., Hasebe, F., Ito, S., 2019. Deracemization and Stereoconversion to Aromatic D -Amino Acid Derivatives with Ancestral L -Amino Acid Oxidase. *ACS Catal.* 9, 10152–10158. <https://doi.org/10.1021/acscatal.9b03418>
- Naser, D., Tarasca, M.V., Siebeneichler, B., Schaefer, A., Deol, H.K., Soule, T.G.B., Almey, J., Kelso, S., Mishra, G.G., Simon, H., Meiering, E.M., 2022. High-Resolution NMR H/D Exchange of Human Superoxide Dismutase Inclusion Bodies Reveals Significant Native Features Despite Structural Heterogeneity. *Angew. Chem. Int. Ed.* 61. <https://doi.org/10.1002/anie.202112645>
- Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* 32, 268–274. <https://doi.org/10.1093/molbev/msu300>
- Oda, M., Yamagami, Y., Inaba, S., Oida, T., Yamamoto, M., Kitajima, S., Kawai, F., 2018. Enzymatic hydrolysis of PET: functional roles of three Ca²⁺ ions bound to a cutinase-like enzyme, Cut190*, and its engineering for improved activity. *Appl. Microbiol. Biotechnol.* 102, 10067–10077. <https://doi.org/10.1007/s00253-018-9374-x>

- Otzen, D.E., Giehm, L., Baptista, R.P., Kristensen, S.R., Melo, E.P., Pedersen, S., 2007. Aggregation as the basis for complex behaviour of cutinase in different denaturants. *Biochim. Biophys. Acta BBA - Proteins Proteomics* 1774, 323–333. <https://doi.org/10.1016/j.bbapap.2006.11.012>
- Pagel, M., Meade, A., Barker, D., 2004. Bayesian Estimation of Ancestral Character States on Phylogenies. *Syst. Biol.* 53, 673–684. <https://doi.org/10.1080/10635150490522232>
- Park, H., Bradley, P., Greisen, P., Liu, Y., Mulligan, V.K., Kim, D.E., Baker, D., DiMaio, F., 2016. Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theory Comput.* 12, 6201–6212. <https://doi.org/10.1021/acs.jctc.6b00819>
- Pauling, L., Zuckerkandl, E., Henriksen, T., Löfstad, R., 1963. Chemical Paleogenetics. Molecular “Restoration Studies” of Extinct Forms of Life. *Acta Chem. Scand.* 17 suppl., 9–16. <https://doi.org/10.3891/acta.chem.scand.17s-0009>
- Penn, O., Privman, E., Ashkenazy, H., Landan, G., Graur, D., Pupko, T., 2010. GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Res.* 38, W23–W28. <https://doi.org/10.1093/nar/gkq443>
- Pereira, J., Simpkin, A.J., Hartmann, M.D., Rigden, D.J., Keegan, R.M., Lupas, A.N., 2021. High-accuracy protein structure prediction in CASP14. *Proteins Struct. Funct. Bioinforma.* 89, 1687–1699. <https://doi.org/10.1002/prot.26171>
- Perez, J.M., Kontur, W.S., Alherech, M., Coplien, J., Karlen, S.D., Stahl, S.S., Donohue, T.J., Noguera, D.R., 2019. Funneling aromatic products of chemically depolymerized lignin into 2-pyrone-4-6-dicarboxylic acid with *Novosphingobium aromaticivorans*. *Green Chem.* 21, 1340–1350. <https://doi.org/10.1039/C8GC03504K>
- Perlmutter, N., Theobald, D., 2020. Department of Biochemistry (Master’s Thesis). Brandeis University, Boston, Massachusetts.
- Pfaff, L., Gao, J., Li, Z., Jäckering, A., Weber, G., Mican, J., Chen, Y., Dong, W., Han, X., Feiler, C.G., Ao, Y.-F., Badenhorst, C.P.S., Bednar, D., Palm, G.J., Lammers, M., Damborsky, J., Strodel, B., Liu, W., Bornscheuer, U.T., Wei, R., 2022. Multiple Substrate Binding Mode-Guided Engineering of a Thermophilic PET Hydrolase. *ACS Catal.* 12, 9790–9800. <https://doi.org/10.1021/acscatal.2c02275>
- Pirillo, V., Orlando, M., Tessaro, D., Pollegioni, L., Molla, G., 2021. An Efficient Protein Evolution Workflow for the Improvement of Bacterial PET Hydrolyzing Enzymes. *Int. J. Mol. Sci.* 23, 264. <https://doi.org/10.3390/ijms23010264>
- PlasticsEurope, 2022. *Plastics - the Facts 2022*.
- Porebski, B.T., Buckle, A.M., 2016. Consensus protein design. *Protein Eng. Des. Sel.* 29, 245–251. <https://doi.org/10.1093/protein/gzw015>
- Priya, A., Dutta, K., Daverey, A., 2022. A comprehensive biotechnological and molecular insight into plastic degradation by microbial community. *J. Chem. Technol. Biotechnol.* 97, 381–390. <https://doi.org/10.1002/jctb.6675>
- Pupko, T., Pe, I., Shamir, R., Graur, D., 2000. A Fast Algorithm for Joint Reconstruction of Ancestral Amino Acid Sequences. *Mol. Biol. Evol.* 17, 890–896. <https://doi.org/10.1093/oxfordjournals.molbev.a026369>
- Puspitasari, N., Lee, C.-K., 2021. Class I hydrophobin fusion with cellulose binding domain for its soluble expression and facile purification. *Int. J. Biol. Macromol.* 193, 38–43. <https://doi.org/10.1016/j.ijbiomac.2021.10.089>

- Qiao, Y., Hu, R., Chen, D., Wang, L., Wang, Z., Yu, H., Fu, Y., Li, C., Dong, Z., Weng, Y.-X., Du, W., 2022. Fluorescence-activated droplet sorting of PET degrading microorganisms. *J. Hazard. Mater.* 424, 127417. <https://doi.org/10.1016/j.jhazmat.2021.127417>
- Ragaert, K., Delva, L., Van Geem, K., 2017. Mechanical and chemical recycling of solid plastic waste. *Waste Manag.* 69, 24–58. <https://doi.org/10.1016/j.wasman.2017.07.044>
- Rambaut, A., 2018. FigTree. <http://tree.bio.ed.ac.uk/software/figtree/>.
- Rennison, A., Winther, J.R., Varrone, C., 2021. Rational Protein Engineering to Increase the Activity and Stability of IsPETase Using the PROSS Algorithm. *Polymers* 13, 3884. <https://doi.org/10.3390/polym13223884>
- Ribitsch, D., Acero, E.H., Greimel, K., Eiteljoerg, I., Trotscha, E., Freddi, G., Schwab, H., Guebitz, G.M., 2012. Characterization of a new cutinase from *Thermobifida alba* for PET-surface hydrolysis. *Biocatal. Biotransformation* 30, 2–9. <https://doi.org/10.3109/10242422.2012.644435>
- Ribitsch, D., Herrero Acero, E., Greimel, K., Dellacher, A., Zitzenbacher, S., Marold, A., Rodriguez, R.D., Steinkellner, G., Gruber, K., Schwab, H., Guebitz, G.M., 2012. A New Esterase from *Thermobifida halotolerans* Hydrolyses Polyethylene Terephthalate (PET) and Polylactic Acid (PLA). *Polymers* 4, 617–629. <https://doi.org/10.3390/polym4010617>
- Ribitsch, D., Heumann, S., Trotscha, E., Herrero Acero, E., Greimel, K., Leber, R., Birner-Gruenberger, R., Deller, S., Eiteljoerg, I., Remler, P., Weber, T., Siegert, P., Maurer, K.-H., Donelli, I., Freddi, G., Schwab, H., Guebitz, G.M., 2011. Hydrolysis of polyethyleneterephthalate by p-nitrobenzylesterase from *Bacillus subtilis*. *Biotechnol. Prog.* 27, 951–960. <https://doi.org/10.1002/btpr.610>
- Risso, V.A., Gavira, J.A., Mejia-Carmona, D.F., Gaucher, E.A., Sanchez-Ruiz, J.M., 2013. Hyperstability and Substrate Promiscuity in Laboratory Resurrections of Precambrian β -Lactamases. *J. Am. Chem. Soc.* 135, 2899–2902. <https://doi.org/10.1021/ja311630a>
- Rodrigues, M.O., Abrantes, N., Gonçalves, F.J.M., Nogueira, H., Marques, J.C., Gonçalves, A.M.M., 2019. Impacts of plastic products used in daily life on the environment and human health: What is known? *Environ. Toxicol. Pharmacol.* 72, 103239. <https://doi.org/10.1016/j.etap.2019.103239>
- Ronkvist, Å.M., Xie, W., Lu, W., Gross, R.A., 2009. Cutinase-Catalyzed Hydrolysis of Poly(ethylene terephthalate). *Macromolecules* 42, 5128–5138. <https://doi.org/10.1021/ma9005318>
- Roth, C., Wei, R., Oeser, T., Then, J., Föllner, C., Zimmermann, W., Sträter, N., 2014. Structural and functional studies on a thermostable polyethylene terephthalate degrading hydrolase from *Thermobifida fusca*. *Appl. Microbiol. Biotechnol.* 98, 7815–7823. <https://doi.org/10.1007/s00253-014-5672-0>
- Sagong, H.-Y., Kim, S., Lee, D., Hong, H., Lee, S.H., Seo, H., Kim, K.-J., 2022. Structural and functional characterization of an auxiliary domain-containing PET hydrolase from Burkholderiales bacterium. *J. Hazard. Mater.* 429, 128267. <https://doi.org/10.1016/j.jhazmat.2022.128267>
- Sagong, H.-Y., Son, H.F., Seo, H., Hong, H., Lee, D., Kim, K.-J., 2021. Implications for the PET decomposition mechanism through similarity and dissimilarity between PETases from *Rhizobacter gummiphilus* and *Ideonella sakaiensis*. *J. Hazard. Mater.* 416, 126075. <https://doi.org/10.1016/j.jhazmat.2021.126075>
- Šali, A., Blundell, T.L., 1993. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* 234, 779–815. <https://doi.org/10.1006/jmbi.1993.1626>

- Sankoff, D., 1975. Minimal Mutation Trees of Sequences. *SIAM J. Appl. Math.* 28, 35–42.
<https://doi.org/10.1137/0128004>
- Schluter, D., Price, T., Mooers, A.Ø., Ludwig, D., 1997. Likelihood of Ancestor States in Adaptive Radiation. *Evolution* 51, 1699–1711. <https://doi.org/10.1111/j.1558-5646.1997.tb05095.x>
- Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., Serrano, L., 2005. The FoldX web server: an online force field. *Nucleic Acids Res.* 33, W382–W388.
<https://doi.org/10.1093/nar/gki387>
- Selke, S.E., 2000. Plastics Recycling and Biodegradable Plastics, in: Harper, C.A. (Ed.), *Modern Plastics Handbook*. McGraw-Hill Education, New York.
- Seo, H., Kim, S., Son, H.F., Sagong, H.-Y., Joo, S., Kim, K.-J., 2019. Production of extracellular PETase from *Ideonella sakaiensis* using sec-dependent signal peptides in *E. coli*. *Biochem. Biophys. Res. Commun.* 508, 250–255.
<https://doi.org/10.1016/j.bbrc.2018.11.087>
- Shi, L., Liu, H., Gao, S., Weng, Y., Zhu, L., 2021. Enhanced Extracellular Production of *Is* PETase in *Escherichia coli* via Engineering of the *pelB* Signal Peptide. *J. Agric. Food Chem.* 69, 2245–2252. <https://doi.org/10.1021/acs.jafc.0c07469>
- Shipton, L., Dauvergne, P., 2022. Health concerns of plastics: energizing the global diffusion of anti-plastic norms. *J. Environ. Plan. Manag.* 65, 2124–2144.
<https://doi.org/10.1080/09640568.2021.1957796>
- Shirke, A.N., White, C., Englaender, J.A., Zwarycz, A., Butterfoss, G.L., Linhardt, R.J., Gross, R.A., 2018. Stabilizing Leaf and Branch Compost Cutinase (LCC) with Glycosylation: Mechanism and Effect on PET Hydrolysis. *Biochemistry* 57, 1190–1200.
<https://doi.org/10.1021/acs.biochem.7b01189>
- Shiryaev, S.A., Cieplak, P., Chernov, A.V., 2021. Salt leverages polyethylene terephthalate hydrolase (PETase) enzymatic activity via the predicted conformational switch (preprint). *Molecular Biology*. <https://doi.org/10.1101/2021.09.23.461413>
- Silva, C., Da, S., Silva, N., Matamá, T., Araújo, R., Martins, M., Chen, S., Chen, J., Wu, J., Casal, M., Cavaco-Paulo, A., 2011. Engineered *Thermobifida fusca* cutinase with increased activity on polyester substrates. *Biotechnol. J.* 6, 1230–1239.
<https://doi.org/10.1002/biot.201000391>
- Skariyachan, S., Setlur, A.S., Naik, S.Y., Naik, A.A., Usharani, M., Vasist, K.S., 2017. Enhanced biodegradation of low and high-density polyethylene by novel bacterial consortia formulated from plastic-contaminated cow dung under thermophilic conditions. *Environ. Sci. Pollut. Res.* 24, 8443–8457. <https://doi.org/10.1007/s11356-017-8537-0>
- Son, H.F., Cho, I.J., Joo, S., Seo, H., Sagong, H.-Y., Choi, S.Y., Lee, S.Y., Kim, K.-J., 2019. Rational Protein Engineering of Thermo-Stable PETase from *Ideonella sakaiensis* for Highly Efficient PET Degradation. *ACS Catal.* 9, 3519–3526.
<https://doi.org/10.1021/acscatal.9b00568>
- Son, H.F., Joo, S., Seo, H., Sagong, H.-Y., Lee, S.H., Hong, H., Kim, K.-J., 2020. Structural bioinformatics-based protein engineering of thermo-stable PETase from *Ideonella sakaiensis*. *Enzyme Microb. Technol.* 141, 109656.
<https://doi.org/10.1016/j.enzmictec.2020.109656>
- Sonnendecker, C., Oeser, J., Richter, P.K., Hille, P., Zhao, Z., Fischer, C., Lippold, H., Blázquez-Sánchez, P., Engelberger, F., Ramírez-Sarmiento, C.A., Oeser, T., Lihanova, Y., Frank, R., Jahnke, H., Billig, S., Abel, B., Sträter, N., Matysik, J., Zimmermann, W.,

2022. Low Carbon Footprint Recycling of Post-Consumer PET Plastic with a Metagenomic Polyester Hydrolase. *ChemSusChem* 15. <https://doi.org/10.1002/cssc.202101062>
- Sormanni, P., Aprile, F.A., Vendruscolo, M., 2015. The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility. *J. Mol. Biol.* 427, 478–490. <https://doi.org/10.1016/j.jmb.2014.09.026>
- Strategy on Zero Plastic Waste, 2018. . Canadian Council of Ministers of the Environment.
- Sulaiman, S., You, D.-J., Kanaya, E., Koga, Y., Kanaya, S., 2014. Crystal Structure and Thermodynamic and Kinetic Stability of Metagenome-Derived LC-Cutinase. *Biochemistry* 53, 1858–1869. <https://doi.org/10.1021/bi401561p>
- Susko, E., Roger, A.J., 2020. On the Use of Information Criteria for Model Selection in Phylogenetics. *Mol. Biol. Evol.* 37, 549–562. <https://doi.org/10.1093/molbev/msz228>
- Tamoor, M., Samak, N.A., Jia, Y., Mushtaq, M.U., Sher, H., Bibi, M., Xing, J., 2021. Potential Use of Microbial Enzymes for the Conversion of Plastic Waste Into Value-Added Products: A Viable Solution. *Front. Microbiol.* 12, 777727. <https://doi.org/10.3389/fmicb.2021.777727>
- Taniguchi, I., Yoshida, S., Hiraga, K., Miyamoto, K., Kimura, Y., Oda, K., 2019. Biodegradation of PET: Current Status and Application Aspects. *ACS Catal.* 9, 4089–4105. <https://doi.org/10.1021/acscatal.8b05171>
- Then, J., Wei, R., Oeser, T., Gerdt, A., Schmidt, J., Barth, M., Zimmermann, W., 2016. 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. <https://doi.org/10.1002/2211-5463.12053>
- Tian, C., Kasavajhala, K., Belfon, K.A.A., Raguette, L., Huang, H., Miguez, A.N., Bickel, J., Wang, Y., Pincay, J., Wu, Q., Simmerling, C., 2020. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* 16, 528–552. <https://doi.org/10.1021/acs.jctc.9b00591>
- Tian, W., Chen, C., Lei, X., Zhao, J., Liang, J., 2018. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* 46, W363–W367. <https://doi.org/10.1093/nar/gky473>
- Tiwari, P., Kaila, P., Guptasarma, P., 2019. Understanding anomalous mobility of proteins on SDS-PAGE with special reference to the highly acidic extracellular domains of human E- and N-cadherins. *ELECTROPHORESIS* 40, 1273–1281. <https://doi.org/10.1002/elps.201800219>
- Tokuriki, N., Stricher, F., Serrano, L., Tawfik, D.S., 2008. How Protein Stability and New Functions Trade Off. *PLoS Comput. Biol.* 4, e1000002. <https://doi.org/10.1371/journal.pcbi.1000002>
- Tournier, V., Topham, C.M., Gilles, A., David, B., Folgoas, C., Moya-Leclair, E., Kamionka, E., Desrousseaux, M.-L., Texier, H., Gavalda, S., Cot, M., Guémard, E., Dalibey, M., Nomme, J., Cioci, G., Barbe, S., Chateau, M., André, I., Duquesne, S., Marty, A., 2020. An engineered PET depolymerase to break down and recycle plastic bottles. *Nature* 580, 216–219. <https://doi.org/10.1038/s41586-020-2149-4>
- Tripathy, A., Sen, P., Su, B., Briscoe, W.H., 2017. Natural and bioinspired nanostructured bactericidal surfaces. *Adv. Colloid Interface Sci.* 248, 85–104. <https://doi.org/10.1016/j.cis.2017.07.030>

- Tsumoto, K., Ejima, D., Kumagai, I., Arakawa, T., 2003. Practical considerations in refolding proteins from inclusion bodies. *Protein Expr. Purif.* 28, 1–8. [https://doi.org/10.1016/S1046-5928\(02\)00641-1](https://doi.org/10.1016/S1046-5928(02)00641-1)
- Wagner-Egea, P., Tosi, V., Wang, P., Grey, C., Zhang, B., Linares-Pastén, J.A., 2021. Assessment of IsPETase-Assisted Depolymerization of Terephthalate Aromatic Polyesters and the Effect of the Thioredoxin Fusion Domain. *Appl. Sci.* 11, 8315. <https://doi.org/10.3390/app11188315>
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303. <https://doi.org/10.1093/nar/gky427>
- Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>
- Wei, R., Oeser, T., Schmidt, J., Meier, R., Barth, M., Then, J., Zimmermann, W., 2016. Engineered bacterial polyester hydrolases efficiently degrade polyethylene terephthalate due to relieved product inhibition: Engineered Polyester Hydrolases. *Biotechnol. Bioeng.* 113, 1658–1665. <https://doi.org/10.1002/bit.25941>
- Wei, R., Oeser, T., Then, J., Kühn, N., Barth, M., Schmidt, J., Zimmermann, W., 2014. Functional characterization and structural modeling of synthetic polyester-degrading hydrolases from *Thermomonospora curvata*. *AMB Express* 4, 44. <https://doi.org/10.1186/s13568-014-0044-9>
- Wei, R., Song, C., Gräsing, D., Schneider, T., Bielytskyi, P., Böttcher, D., Matysik, J., Bornscheuer, U.T., Zimmermann, W., 2019. Conformational fitting of a flexible oligomeric substrate does not explain the enzymatic PET degradation. *Nat. Commun.* 10, 5581. <https://doi.org/10.1038/s41467-019-13492-9>
- Weigert, S., Perez-Garcia, P., Gisdon, F.J., Gagsteiger, A., Schweinshaut, K., Ullmann, G.M., Chow, J., Streit, W.R., Höcker, B., 2022. Investigation of the halophilic PET hydrolase PET6 from *Vibrio gazogenes*. *Protein Sci.* 31, e4500. <https://doi.org/10.1002/pro.4500>
- Whitfield, J.H., Zhang, W.H., Herde, M.K., Clifton, B.E., Radziejewski, J., Janovjak, H., Henneberger, C., Jackson, C.J., 2015. Construction of a robust and sensitive arginine biosensor through ancestral protein reconstruction: cpFLIPR: Improved Biosensor for L - Arginine. *Protein Sci.* 24, 1412–1422. <https://doi.org/10.1002/pro.2721>
- Wiesinger, H., Wang, Z., Hellweg, S., 2021. Deep Dive into Plastic Monomers, Additives, and Processing Aids. *Environ. Sci. Technol.* 55, 9339–9351. <https://doi.org/10.1021/acs.est.1c00976>
- Wilding, M., Peat, T.S., Kalyaanamoorthy, S., Newman, J., Scott, C., Jermiin, L.S., 2017. Reverse engineering: transaminase biocatalyst development using ancestral sequence reconstruction. *Green Chem.* 19, 5375–5380. <https://doi.org/10.1039/C7GC02343J>
- Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., Jain, S., Lewis, S.M., Arendall, W.B., Snoeyink, J., Adams, P.D., Lovell, S.C., Richardson, J.S., Richardson, D.C., 2018. MolProbity: More and better reference data for improved all-atom structure validation: PROTEIN SCIENCE.ORG. *Protein Sci.* 27, 293–315. <https://doi.org/10.1002/pro.3330>

- Williams, P.D., Pollock, D.D., Blackburne, B.P., Goldstein, R.A., 2006. Assessing the Accuracy of Ancestral Protein Reconstruction Methods. *PLoS Comput. Biol.* 2, e69. <https://doi.org/10.1371/journal.pcbi.0020069>
- Wright, S.L., Kelly, F.J., 2017. Plastic and Human Health: A Micro Issue? *Environ. Sci. Technol.* 51, 6634–6647. <https://doi.org/10.1021/acs.est.7b00423>
- Xi, X., Ni, K., Hao, H., Shang, Y., Zhao, B., Qian, Z., 2021. Secretory expression in *Bacillus subtilis* and biochemical characterization of a highly thermostable polyethylene terephthalate hydrolase from bacterium HR29. *Enzyme Microb. Technol.* 143, 109715. <https://doi.org/10.1016/j.enzmictec.2020.109715>
- Yang, L.-L., Tang, S.-K., Zhang, Y.-Q., Zhi, X.-Y., Wang, D., Xu, L.-H., Li, W.-J., 2008. *Thermobifida halotolerans* sp. nov., isolated from a salt mine sample, and emended description of the genus *Thermobifida*. *Int. J. Syst. Evol. Microbiol.* 58, 1821–1825. <https://doi.org/10.1099/ijs.0.65732-0>
- Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., Toyohara, K., Miyamoto, K., Kimura, Y., Oda, K., 2016. A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science* 351, 1196–1199. <https://doi.org/10.1126/science.aad6359>
- Zakas, P.M., Brown, H.C., Knight, K., Meeks, S.L., Spencer, H.T., Gaucher, E.A., Doering, C.B., 2017. Enhancing the pharmaceutical properties of protein drugs by ancestral sequence reconstruction. *Nat. Biotechnol.* 35, 35–37. <https://doi.org/10.1038/nbt.3677>
- Zalasiewicz, J., Waters, C.N., Ivar do Sul, J.A., Corcoran, P.L., Barnosky, A.D., Cearreta, A., Edgeworth, M., Gałuszka, A., Jeandel, C., Leinfelder, R., McNeill, J.R., Steffen, W., Summerhayes, C., Wagleich, M., Williams, M., Wolfe, A.P., Yonan, Y., 2016. The geological cycle of plastics and their use as a stratigraphic indicator of the Anthropocene. *Anthropocene* 13, 4–17. <https://doi.org/10.1016/j.ancene.2016.01.002>
- Zeller, M., Netsch, N., Richter, F., Leibold, H., Stapf, D., 2021. Chemical Recycling of Mixed Plastic Wastes by Pyrolysis – Pilot Scale Investigations. *Chem. Ing. Tech.* 93, 1763–1770. <https://doi.org/10.1002/cite.202100102>
- Zhang, H., Perez-Garcia, P., Dierkes, R.F., Applegate, V., Schumacher, J., Chibani, C.M., Sternagel, S., Preuss, L., Weigert, S., Schmeisser, C., Danso, D., Pleiss, J., Almeida, A., Höcker, B., Hallam, S.J., Schmitz, R.A., Smits, S.H.J., Chow, J., Streit, W.R., 2022. The *Bacteroidetes* *Aequorivita* sp. and *Kaistella jeonii* Produce Promiscuous Esterases With PET-Hydrolyzing Activity. *Front. Microbiol.* 12, 803896. <https://doi.org/10.3389/fmicb.2021.803896>
- Zhong-Johnson, E.Z.L., Voigt, C.A., Sinskey, A.J., 2021. An absorbance method for analysis of enzymatic degradation kinetics of poly(ethylene terephthalate) films. *Sci. Rep.* 11, 928. <https://doi.org/10.1038/s41598-020-79031-5>
- Zimmermann, M.T., Jernigan, R.L., 2012. Protein Loop Dynamics Are Complex and Depend on the Motions of the Whole Protein. *Entropy* 14, 687–700. <https://doi.org/10.3390/e14040687>
- Zimmermann, W., WEI, R., HILLE, P., Oeser, T., SCHMIDT, J., 2019. New polypeptides having a polyester degrading activity and uses thereof.

Appendices

A. Engineering of Known PET Hydrolases

Table A1 Literature Survey of PET Hydrolase Variants. The below table is an attempt at synthesizing variants of PET hydrolases that have been described in literature. Here, the enzyme describes the starting point, while the change column describes the mutation or alteration that has occurred. Note that all residues are numbered according to cSer (in the enzyme column). Stability is listed both as the experimentally measured T_m (unless otherwise specified), as well as the degree of change from the wild-type enzyme. Activity is reported simply as an increase or decrease against a particular substrate, as there are many external factors that can influence activity (buffer, pH, additives, temperature, substrate, etc.). Finally, the strategy refers to the basis for the alteration or mutation chosen. For succinctness, there are a number of abbreviations in the table: Hc PET = high crystallinity, Lc PET = low crystallinity, npPET = nano-plastic PET, pNPA = p-nitrophenol acetate, mpPET = microplastic PET, pfPET = preform PET, pw=powder PET, fibPET = PET fibers, CBM66 = carbohydrate binding module.

Enzyme	Change	Stability		Activity		Strategy	Reference
Tf. Cut II (cSer=130)	+ -N(CH ₃) ₃ ⁺	-	-	Substantially Increased	lcPET	Alter enzyme-PET surface binding	(Furukawa, Kawakami, ... Miyamoto, 2019)
	-OSO ₃ ⁻	-	-	Increased	lcPET	Alter enzyme-PET surface binding	(Furukawa, Kawakami, ... Miyamoto, 2019)
	G62A	-	-	Improved	lcPET	Altering non-conserved sites around active site	(Furukawa, Kawakami, ... Miyamoto, 2019)
	G62A	-	-	Improved	lcPET	Match LCC substrate binding residues Reduces product inhibition	(Wei, Oeser, ... Zimmermann, 2016)
	T63D	-	-	Similar	lcPET	Match LCC substrate binding residues	(Wei, Oeser, ... Zimmermann, 2016)
	H129W	-	-	Improved	lcPET	Altering non-conserved sites around active site	(Furukawa, Kawakami, ... Miyamoto, 2019)
	H129A	-	-	Substantially Decreased	lcPET	Altering non-conserved sites around active site	(Furukawa, Kawakami, ... Miyamoto, 2019)
	I178V	-	-	Similar	lcPET	Match LCC substrate binding residues	(Wei, Oeser, ... Zimmermann, 2016)

F209S	-	-	Increased	lcPET	Altering non-conserved sites around active site	(Furukawa, Kawakami, ... Miyamoto, 2019)
F209A	-	-	Increased	lcPET	Altering non-conserved sites around active site	(Furukawa, Kawakami, ... Miyamoto, 2019)
I213S	-	-	Similar	lcPET	Match LCC substrate binding residues	(Wei, Oeser, ... Zimmermann, 2016)
H184S F188I	-	-	Increased	PET granules	Match Is. PETase unique residues	(Chen, Han, ... Guo, 2021)
G62A T63D	-	-	Increased (Similar to G62A)	lcPET	Combine active G62A mutant with other mutations to match LCC residues	(Wei, Oeser, ... Zimmermann, 2016)
G62A I178V	-	-	Increased (Similar to G62A)	lcPET	Combine active G62A mutant with other mutations to match LCC residues	(Wei, Oeser, ... Zimmermann, 2016)
G62A I213S	-	-	Increased (Similar to G62A)	lcPET	Combine active G62A mutant with other mutations to match LCC residues	(Wei, Oeser, ... Zimmermann, 2016)
G62A T63D I178V	-	-	Increased (Similar to G62A)	lcPET	Combine active G62A mutant with other mutations to match LCC residues	(Wei, Oeser, ... Zimmermann, 2016)
G62A T63D I213S	-	-	Increased (Similar to G62A)	lcPET	Combine active G62A mutant with other mutations to match LCC residues	(Wei, Oeser, ... Zimmermann, 2016)
G62A I178V I213S	-	-	Increased (Similar to G62A)	lcPET	Combine active G62A mutant with other mutations to match LCC residues	(Wei, Oeser, ... Zimmermann, 2016)
G62A T63D I178V I213S	-	-	Increased (Similar to G62A)	lcPET	Combine active G62A mutant with other mutations to match LCC residues	(Wei, Oeser, ... Zimmermann, 2016)
G62A H129W	-	-	Increased	lcPET	Combining active single mutations	(Furukawa, Kawakami, ... Miyamoto, 2019)
G62A F209S	-	-	Increased	lcPET	Combining active single mutations	(Furukawa, Kawakami, ... Miyamoto, 2019)

	G62A F209A	-	-	Increased	lcPET	Combining active single mutations	(Furukawa, Kawakami, ... Miyamoto, 2019)
	H129W F209S	-	-	Similar	lcPET	Combining active single mutations	(Furukawa, Kawakami, ... Miyamoto, 2019)
	H129W F209A	-	-	Substantially Decreased	lcPET	Combining active single mutations	(Furukawa, Kawakami, ... Miyamoto, 2019)
	D174C-D204C	69.8	N.C.	Similar	lcPET	Disulfide addition to replace Ca+2 binding site	(Then, Wei, ... Zimmermann, 2016)
	D174C-D253C	88.3	+18.5	Decreased	lcPET	Disulfide addition to replace Ca+2 binding site	(Then, Wei, ... Zimmermann, 2016)
	D204C-E253C	92.8	+23	Increased	lcPET	Disulfide addition to replace Ca+2 binding site	(Then, Wei, ... Zimmermann, 2016)
	D204C-E253C D174L	Lower than D204C-E253C		Increased	lcPET	Non-polar residue addition to remove destabilization at Ca+2 binding site	(Then, Wei, ... Zimmermann, 2016)
	D204C-E253C D174N	92.8	+23	Increased	lcPET	Reduction of negative charge to remove destabilization at Ca+2 binding site	(Then, Wei, ... Zimmermann, 2016)
	D204C-E253C D174A	92.8	+23	Increased	lcPET	Steric hinderance removal to remove destabilization at Ca+2 binding site	(Then, Wei, ... Zimmermann, 2016)
	D204C-E253C D174R	92.8	+23	Increased	lcPET	Reduction of negative charge to remove destabilization at Ca+2 binding site	(Then, Wei, ... Zimmermann, 2016)
	D204C-E253C D174R-G205D	94.6	+24.8	Increased	lcPET	Introduction of salt bridge to mimic Ca+2 attachment of loops	(Then, Wei, ... Zimmermann, 2016)
Tfu_0883 (cSer = 130)	I178A	-	-	Increased	PET fabric	Replacement of bulky side chain with smaller one to improve PET binding	(Silva, Da, ... Cavaco-Paulo, 2011)
	Q92A T61A	-	-	Increased	PET fabric	Replacement of bulky side chain with smaller one to improve PET binding	(Silva, Da, ... Cavaco-Paulo, 2011)
The_CutI (cSer = 131)	N29D N49D N161D	-	-	Increased	lcPET	Knock out glycosylation site for expression in <i>P. pastoris</i>	(Gamerith, Vastano, ... Guebitz, 2017)
	S31A T51A S163A	-	-	Increased	lcPET	Knock out glycosylation site for expression in <i>P. pastoris</i>	(Gamerith, Vastano, ... Guebitz, 2017)
The_Cut II (cSer = 131)	R19S	-	-	Similar	PET film	Exchange of The_Cut I residue into The_Cut II	(Herrero Acero, Ribitsch, ... Guebitz, 2013)

R29N	-	-	Similar	PET film	Exchange of Thc_Cut I residue into Thc_Cut II	(Herrero Acero, Ribitsch, ... Guebitz, 2013)
A30V	-	-	Increased	PET film	Exchange of Thc_Cut I residue into Thc_Cut II	(Herrero Acero, Ribitsch, ... Guebitz, 2013)
Q65E	-	-	No Activity	PET film	Exchange of Thc_Cut I residue into Thc_Cut II	(Herrero Acero, Ribitsch, ... Guebitz, 2013)
L183A	-	-	Similar	PET film	Exchange of Thc_Cut I residue into Thc_Cut II	(Herrero Acero, Ribitsch, ... Guebitz, 2013)
R187K	-	-	Decreased	PET film	Exchange of Thc_Cut I residue into Thc_Cut II	(Herrero Acero, Ribitsch, ... Guebitz, 2013)
R29N A30V	-	-	Increased	PET film	Combination of single mutations	(Herrero Acero, Ribitsch, ... Guebitz, 2013)
R19S R29N A30V	-	-	Increased	PET film	Combination of single mutations	(Herrero Acero, Ribitsch, ... Guebitz, 2013)
Cut190 (cSer=130)	F60Y	-	-	-	F in this position is thought to seal the active site. Tyr, allows active site formation.	(Miyakawa, Mizushima, ... Tanokura, 2015)
	S130A	-	-	Inactive	-	Alanine mutation of cSer (Kawabata, Oda, and Kawai, 2017)
	S180P	76.5 (with Ca²⁺)	+5.9	Increased	lcPET	Consensus matching (Kawai, Oda, ... Tanokura, 2014)
	S180P R182S	76.8 (with Ca²⁺)	+6.2	Increased	lcPET	Break salt bridge, increase flexibility of cAsp loop (Kawai, Oda, ... Tanokura, 2014)
Cut190* (S180P) (R182S)	T216K	76.2 (with Ca²⁺)	+5.6	Increased	lcPET	Make salt bridge at end of cHis loop (Kawai, Oda, ... Tanokura, 2014)
	+Ser between res 77, 78	55.8	-0.1	Increased	PBSA	Reduced Ca ²⁺ dependence for activity, but higher insoluble expression (Oda, Yamagami, ... Kawai, 2018)
	F60A	-	-	Decreased	PBSA	Mutated based on docking to PET substrate (Kawabata, Oda, and Kawai, 2017)
	F60Y	-	-	Decreased	PBSA	Mutated based on docking to PET substrate (Kawabata, Oda, and Kawai, 2017)

T61A	-	-	Similar	PBSA	Mutated based on docking to PET substrate	(Kawabata, Oda, and Kawai, 2017)
Q64E	59.4	+3.5	Similar	PBSA	Modification of surface Gln and Asn	(Oda, Yamagami, ... Kawai, 2018)
Q64A	-	-	-	-	Poor expression Modification of surface Gln and Asn	(Oda, Yamagami, ... Kawai, 2018)
S66A	-	-	Similar	PBSA	Mutated based on docking to PET substrate	(Kawabata, Oda, and Kawai, 2017)
Q77H	58.6	+2.7	Increased	PBSA	Mutated based on docking to PET substrate	(Oda, Yamagami, ... Kawai, 2018)
N87D	60.4	+4.5	Similar	PBSA	Modification of surface Gln and Asn	(Oda, Yamagami, ... Kawai, 2018)
Q92A	57.3	+1.4	Increased	PBSA mcPET	Mutated based on docking to PET substrate	(Kawabata, Oda, and Kawai, 2017)
Q92L	-	-	No Activity	PBSA	Mutated based on docking to PET substrate	(Kawabata, Oda, and Kawai, 2017)
Q92D	-	-	Slightly Decreased	PBSA	Mutated based on docking to PET substrate	(Kawabata, Oda, and Kawai, 2017)
Q95E	60.7	+4.8	Increased	PBSA	Modification of surface Gln and Asn	(Oda, Yamagami, ... Kawai, 2018)
Q95A	-	-	-	-	Poor expression Modification of surface Gln and Asn	(Oda, Yamagami, ... Kawai, 2018)
N212A	-	-	Similar	PBSA	Mutated based on docking to PET substrate	(Kawabata, Oda, and Kawai, 2017)
N122D	56	-0.1	Similar	PBSA	Modification of surface Gln and Asn	(Oda, Yamagami, ... Kawai, 2018)
N122A	56	-0.1	Increased	PBSA	Modification of surface Gln and Asn	(Oda, Yamagami, ... Kawai, 2018)
H129A	-	-	No Activity	PBSA	Mutated based on docking to PET substrate	(Kawabata, Oda, and Kawai, 2017)
M131A	-	-	No Activity	PBSA	Mutated based on docking to PET substrate	(Kawabata, Oda, and Kawai, 2017)
E138R	53.9	-2	Increased	PBSA	Interferes with Ca ²⁺ binding Considered to form salt bridge	(Oda, Yamagami, ... Kawai, 2018)
W155A			Decreased	PBSA	Mutated based on docking to PET substrate	(Kawabata, Oda, and Kawai, 2017)
N156H	62.4	+6.5	Increased	PBSA	Mutated based on docking to PET substrate	(Oda, Yamagami, ... Kawai, 2018)
Q163E	58.6	+2.7	Increased	PBSA	Mutated based on docking to PET substrate	(Oda, Yamagami, ... Kawai, 2018)

I178A	-	-	Increased Decreased	PBSA mcPET	Mutated based on docking to PET substrate	(Kawabata, Oda, and Kawai, 2017)	
F209A	-	-	-	-	Poor expression Mutated based on docking to PET substrate	(Kawabata, Oda, and Kawai, 2017)	
I178A T177V	-	-	-	-	Poor expression Mutated based on docking to PET substrate	(Kawabata, Oda, and Kawai, 2017)	
I178A Q92A			Increased	PBSA	Mutated based on docking to PET substrate	(Kawabata, Oda, and Kawai, 2017)	
I178A N212A			Increased	PBSA	Mutated based on docking to PET substrate	(Kawabata, Oda, and Kawai, 2017)	
Q92A Q95E	57.9	+2	Decreased	PBSA	Combination of above mutations	(Oda, Yamagami, ... Kawai, 2018)	
D204C-E250C	79.0	+23.1	Increased	PBSA	Replacement of known Ca ²⁺ binding site with disulfide bond	(Oda, Yamagami, ... Kawai, 2018)	
Q92A D204C-E250C	79.2	+23.3	Increased	PBSA	Combination of above mutations	(Oda, Yamagami, ... Kawai, 2018)	
Q92A E174R D204C-E250C	78.3	+22.4	Increased	PBSA	Combination of above mutations	(Oda, Yamagami, ... Kawai, 2018)	
Q92A E174R-G205D D204C-E250C	75.8	+19.9	Increased	PBSA	Combination of above mutations	(Oda, Yamagami, ... Kawai, 2018)	
Q92A D204-E250C N156H	85.7	+29.8	Increased	PBSA	Combination of above mutations	(Oda, Yamagami, ... Kawai, 2018)	
Q92A D204-E250C E174R-G205D N156H	79.8	+23.9	Increased	PBSA	Combination of above mutations	(Oda, Yamagami, ... Kawai, 2018)	
PE-H (cSer=135)	G217S	39.7	-11.1	Substantially Decreased	lcPET, hcPET	Match Is. PETase cHis loop	(Bollinger, Thies, ... Jaeger, 2020)
	S219N	49.5	-1.3	Decreased lcPET, Increased hcPET	lcPET, hcPET	Match Is. PETase cHis loop	(Bollinger, Thies, ... Jaeger, 2020)
	I220S	48.5	-2.3	Decreased	lcPET, hcPET	Match Is. PETase cHis loop	(Bollinger, Thies, ... Jaeger, 2020)

	Y221N	43.8	-7	Decreased	lcPET, hcPET	Match Is. PETase cHis loop	(Bollinger, Thies, ... Jaeger, 2020)
	N222Q	45.2	-5.6	Decreased	lcPET, hcPET	Match Is. PETase cHis loop	(Bollinger, Thies, ... Jaeger, 2020)
	Y213S	49.8	-1	Increased lcPET, hcPET	lcPET, hcPET	Known Improving Mutation, Improves substrate accessibility	(Bollinger, Thies, ... Jaeger, 2020)
	G217S S219N I220S Y221N N222Q	39.4	-11.4	Substantially Decreased	lcPET, hcPET	Match Is. PETase cHis loop	(Bollinger, Thies, ... Jaeger, 2020)
Rg. PETase (cSer=132)	H185S	-	-	Substantially decreased	lcPET	Match Is. PETase stabilization of 154-166 loop (near active site)	(Sagong, Son, ... Kim, 2021)
	F189I	-	-	Decreased	lcPET	Match Is. PETase stabilization of 154-166 loop (near active site)	(Sagong, Son, ... Kim, 2021)
	Q192S	-	-	Decreased	lcPET	Match Is. PETase stabilization of 154-166 loop (near active site)	(Sagong, Son, ... Kim, 2021)
	K143A	46	-2	Increased	lcPET	Match Is. PETase stabilization of 154-166 loop (near active site)	(Sagong, Son, ... Kim, 2021)
	E160A	47	-1	Decreased	lcPET	Match Is. PETase stabilization of 154-166 loop (near active site)	(Sagong, Son, ... Kim, 2021)
	R19A	-	-	Similar	lcPET	Explore charge-effect of surface exposed residues	(Sagong, Son, ... Kim, 2021)
	R19E	-	-	Similar	lcPET	Explore charge-effect of surface exposed residues	(Sagong, Son, ... Kim, 2021)
	R31A	-	-	Similar	lcPET	Explore charge-effect of surface exposed residues	(Sagong, Son, ... Kim, 2021)
	R31E	-	-	Similar	lcPET	Explore charge-effect of surface exposed residues	(Sagong, Son, ... Kim, 2021)
	R104A	-	-	Increased	lcPET	Explore charge-effect of surface exposed residues	(Sagong, Son, ... Kim, 2021)
	R104E	-	-	Increased	lcPET	Explore charge-effect of surface exposed residues	(Sagong, Son, ... Kim, 2021)
	R195A	-	-	Similar	lcPET	Explore charge-effect of surface exposed residues	(Sagong, Son, ... Kim, 2021)
	R195E	-	-	Increased	lcPET	Explore charge-effect of surface exposed residues	(Sagong, Son, ... Kim, 2021)
	R62A	-	-	Decreased	lcPET	Explore charge-effect of surface exposed residues	(Sagong, Son, ... Kim, 2021)

	R62E	-	-	Similar	lcPET	Explore charge-effect of surface exposed residues	(Sagong, Son, ... Kim, 2021)
	R253A	-	-	Increased	lcPET	Explore charge-effect of surface exposed residues	(Sagong, Son, ... Kim, 2021)
	R253E	-	-	Increased	lcPET	Explore charge-effect of surface exposed residues	(Sagong, Son, ... Kim, 2021)
	K143A E160A	46	-2	Decreased	lcPET	Abolish salt bridge between Lys and Glu	(Sagong, Son, ... Kim, 2021)
Is. PETase (cSer=131)	+ C _n ⁻ OSO ₃ ⁻	-	-	Improved activity (n=14 most active)	lcPET	Alter enzyme-PET surface binding	(Furukawa, Kawakami, ... Miyamoto, 2018)
	+ C _n -SO ₃ ⁻	-	-	Improved activity (n = 13 most active)	lcPET	Alter enzyme-PET surface binding	(Furukawa, Kawakami, ... Miyamoto, 2018)
	+ C _n -COO ⁻	-	-	Improved activity (n=13 most active)	lcPET	Alter enzyme-PET surface binding	(Furukawa, Kawakami, ... Miyamoto, 2018)
	+ C ₁₂ -N(CH ₃) ₃ ⁺	-	-	Decreased activity substantially	lcPET	Alter enzyme-PET surface binding	(Furukawa, Kawakami, ... Miyamoto, 2018)
	+ RoIA Pre-treatment	-	-	Increased	fibPET hcPET	Improve Is. PETase – surface binding through hydrophobin treatment	(Puspitasari and Lee, 2021)
	+ HGFI Pre-treatment	-	-	Increased	fibPET hcPET	Improve Is. PETase – surface binding through hydrophobin treatment	(Puspitasari and Lee, 2021)
	+ 5 M NaCl (seawater ~ 0.47 M NaCl)	-	-	Increased	pNPA	Mediates conformational changes in cHis and cAsp loop	(Shiryaev, Cieplak, and Chernov, 2021)
	+ Tert-butyl methacrylate modification	Similar	-	Increased	PET film	Chemical modification can improve catalytic activity	(Chen, Quan, ... Sun, 2021)
	+Hydroxyethyl methacrylate modification	Similar	-	Increased	PET film	Chemical modification can improve catalytic activity	(Chen, Quan, ... Sun, 2021)
	+ 2-(dimethylamino) ethyl methacrylate	Similar	-	Increased	PET film	Chemical modification can improve catalytic activity	(Chen, Quan, ... Sun, 2021)

modification + methacrylic acid modification	Similar	-	Increased	PET film	Chemical modification can improve catalytic activity	(Chen, Quan, ... Sun, 2021)
+poly (EK) extension N=5	Improved	-	Increased	PET film	Abundant Glu and Lys found on proteins that exist in challenging conditions: EK tag can protect attached polypeptide	(Chen, Hu, ... Sun, 2021)
+poly (EK) extension N=10	Improved	-	Increased	PET film	Abundant Glu and Lys found on proteins that exist in challenging conditions: EK tag can protect attached polypeptide	(Chen, Hu, ... Sun, 2021)
+poly (EK) extension N=30	Improved	-	Increased	PET film	Abundant Glu and Lys found on proteins that exist in challenging conditions: EK tag can protect attached polypeptide	(Chen, Hu, ... Sun, 2021)
Immobilization to (Cu ₃ (PO ₄) ₂) nanoparticles	Improved	-	Increased	lcPET	Immobilized proteins have demonstrated superior performance for past applications	(Jia, Samak, ... Xing, 2021)
R5E	-	-	Similar	lcPET	Increase negative surface charge to simulate effects of anionic surfactants	(Furukawa, Kawakami, ... Miyamoto, 2018)
R24E	-	-	Similar	lcPET	Increase negative surface charge to simulate effects of anionic surfactants	(Furukawa, Kawakami, ... Miyamoto, 2018)
A36G	51	N.C.	-	-	Premuse strategy	(Meng, Yang, ... Tian, 2021)
T48E	47.5	+1.5	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
G57E	42.0	N.C.	Substantially Decreased	lcPET	Determine role of residues adjacent to active site	(Guo, Vanga, ... Syrén, 2022)
Y58F	-	-	Similar	PET bottle	Structural bioinformatic hotspots for engineering	(Son, Joo, ... Kim, 2020)
Y58A	-	-	Increased	PET bottle	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
Y58A	-	-	Decreased	PET film	Verify role of residues that may be involved in catalytic mechanism by mutation.	(Han, Liu, ... Guo, 2017)
Y58A	-	-	Decreased	PET film	Verify role of residues located in substrate binding site	(Joo, Cho, ... Kim, 2018)
Y58H	42.0	N.C.	Substantially Decreased	lcPET	Determine role of residues adjacent to active site	(Guo, Vanga, ... Syrén, 2022)

Y58E	54.0	+12	Substantially Decreased	lcPET	Determine role of residues adjacent to active site	(Guo, Vanga, ... Syrén, 2022)
T59A	-	-	Slightly Decreased	PET film	Verify role of residues that may be involved in catalytic mechanism by mutation.	(Han, Liu, ... Guo, 2017)
T59L	-	-	Substantially Decreased	PET bottle	Structural bioinformatic hotspots for engineering	(Son, Joo, ... Kim, 2020)
T59H	42.0	N.C.	Decreased	lcPET	Determine role of residues adjacent to active site	(Guo, Vanga, ... Syrén, 2022)
A60S	42.0	N.C.	Similar	lcPET	Determine role of residues adjacent to active site	(Guo, Vanga, ... Syrén, 2022)
A60H	42.0	N.C.	No Activity	lcPET	Determine role of residues adjacent to active site	(Guo, Vanga, ... Syrén, 2022)
R61E	-	-	Substantially Decreased	lcPET	Increase negative surface charge to simulate effects of anionic surfactants	(Furukawa, Kawakami, ... Miyamoto, 2018)
R61A	-	-	Increased	PET film	Improve substrate binding site by replacing large residue with small, hydrophobic one	(Ma, Yao, ... Yuan, 2018)
R61S	-	-	Decreased	PET bottle	Structural bioinformatic hotspots for engineering	(Son, Joo, ... Kim, 2020)
K66E	-	-	Decreased	lcPET	Increase negative surface charge to simulate effects of anionic surfactants	(Furukawa, Kawakami, ... Miyamoto, 2018)
K66A	48.5	+2.5	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
W68L	-	-	No Activity	PET bottle	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
L88F	49.0	+3	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
L88F	-	-	Increased	PET film	Improve substrate binding by mutation to hydrophobic residue. Has affect on Y58, which may improve activity.	(Ma, Yao, ... Yuan, 2018)
Q90Y	50.5	+4.5	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
Q90A	-	-	Decreased	PET bottle	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
S92D	47.35	-1.46	Slightly Increased	PET film	Potential site for H-bond addition	(Son, Cho, ... Kim, 2019)

S92E	47.68	-1.13	Slightly Increased	PET film	Potential site for H-bond addition	(Son, Cho, ... Kim, 2019)
S92E	50.2	+1.5	-	-	MutCompute (Machine Learning)	(Lu, Diaz, ... Alper, 2022)
R94A	-	-	Decreased	PET bottle	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
S96R	43.1	+0.6	-	-	PROSS strategy	(Rennison, Winther, and Varrone, 2021)
Q98L	43	-8	-	-	Premuse strategy	(Meng, Yang, ... Tian, 2021)
R103D	50	-1	-	-	Premuse strategy	(Meng, Yang, ... Tian, 2021)
V105L	50	-1	-	-	Premuse strategy	(Meng, Yang, ... Tian, 2021)
S107E	43.6	+1.2	Increased	lcPET	PROSS strategy	(Rennison, Winther, and Varrone, 2021)
G110S	50	-1	-	-	Premuse strategy	(Meng, Yang, ... Tian, 2021)
T111D	48.0	+2	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
T111D	49.8	+1.1	-	-	MutCompute (Machine Learning)	(Lu, Diaz, ... Alper, 2022)
K119W	48.0	+2	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
M125L	49	-2	-	-	Premuse strategy	(Meng, Yang, ... Tian, 2021)
W130H	55.0	+9	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
W130H	-	-	Increased (pH 7.5, 30 °C)	PET bottle, BHET	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
W130H	57.6	+6	-	-	Premuse strategy	(Meng, Yang, ... Tian, 2021)
W130H	-	-	Decreased (pH 9, 30 °C)	PET film	Verify role of residues that may be involved in catalytic mechanism by mutation.	(Han, Liu, ... Guo, 2017)
W130H	-	-	Decreased (pH 9, 30 °C)	PET film	Verify role of residues located in substrate binding site	(Joo, Cho, ... Kim, 2018)
W130A	-	-	Increased	PET bottle	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)

W130A	-	-	Decreased	PET film	Verify role of residues that may be involved in catalytic mechanism by mutation.	(Han, Liu, ... Guo, 2017)
W130A	-	-	Decreased	PET film	Verify role of residues located in substrate binding site	(Joo, Cho, ... Kim, 2018)
S131A (cSer)	-	-	No Activity	-	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
S131A (cSer)	-	-	No Activity	PET film	Verify role of residues that may be involved in catalytic mechanism by mutation.	(Han, Liu, ... Guo, 2017)
S131A (cSer)	-	-	No Activity	PET film	Verify role of residues located in substrate binding site	(Joo, Cho, ... Kim, 2018)
M132Q	42.0	N.C.	Substantially Decreased	lcPET	Determine role of residues adjacent to active site	(Guo, Vanga, ... Syrén, 2022)
M132A	-	-	Decreased	PET bottle	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
M132A	-	-	Decreased	PET film	Verify role of residues that may be involved in catalytic mechanism by mutation.	(Han, Liu, ... Guo, 2017)
M132A	-	-	Decreased	PET film	Verify role of residues located in substrate binding site	(Joo, Cho, ... Kim, 2018)
G136A	47.5	+1.5	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
S137T	48.0	+2	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
I139R	54.0	+8	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
S140A	48	-3	-	-	Premuse strategy	(Meng, Yang, ... Tian, 2021)
A151I	48.0	+2	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
P152A	52.5	+6.5	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
P152A	49.25	+0.5	Decreased	PET film	Attempt to restore continuity of β -sheet	(Son, Cho, ... Kim, 2019)
P152G	43.79	-5.02	Substantially Decreased	PET film	Attempt to restore continuity of β -sheet	(Son, Cho, ... Kim, 2019)
P152S	45.92	-2.89	Substantially Decreased	PET film	Attempt to restore continuity of β -sheet	(Son, Cho, ... Kim, 2019)

Q153L	-	-	Similar	PET bottle	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
W156A	-	-	Substantially Decreased	hcPET	Determine role of highly conserved W	(Austin, Allen, ... Beckham, 2018)
W156A	-	-	Decreased	PET bottle	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
W156A	-	-	Substantially Decreased	PET film	Verify role of residues that may be involved in catalytic mechanism by mutation.	(Han, Liu, ... Guo, 2017)
W156A	-	-	Decreased	PET film	Verify role of residues located in substrate binding site	(Joo, Cho, ... Kim, 2018)
W156T	42.0	N.C.	Substantially Decreased	lcPET	Determine role of residues adjacent to active site	(Guo, Vanga, ... Syrén, 2022)
D157H	53.0	+7	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
D157H	53.16	+4.35	Similar	PET film	Stabilizes hydrophobic loop by removing polar Asp	(Son, Cho, ... Kim, 2019)
D157F	51.90	+3.09	Similar	PET film	Stabilizes hydrophobic loop by removing polar Asp	(Son, Cho, ... Kim, 2019)
D157I	52.70	+3.89	Substantially Decreased	PET film	Stabilizes hydrophobic loop by removing polar Asp	(Son, Cho, ... Kim, 2019)
D157L	51.67	+2.86	Substantially Decreased	PET film	Stabilizes hydrophobic loop by removing polar Asp	(Son, Cho, ... Kim, 2019)
D157V	53.46	+4.65	Substantially Decreased	PET film	Stabilizes hydrophobic loop by removing polar Asp	(Son, Cho, ... Kim, 2019)
S158W	49.5	+3.5	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
S159Q	47.5	+1.5	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
C174A	-	-	No Activity	pNPA	Determine role of active site disulfide by alanine substitution	(Fecker, Galaz-Davison, ... Ramírez-Sarmiento, 2018)
C174A	-	-	Substantially Decreased	PET bottle	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
C174S	-	-	No Activity	PET film	Verify role of residues that may be involved in catalytic mechanism by mutation.	(Han, Liu, ... Guo, 2017)
D177A (cAsp)	-	-	No Activity	-	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)

D177A (cAsp)	-	-	No Activity	PET film	Verify role of residues located in substrate binding site	(Joo, Cho, ... Kim, 2018)
S178T	-	-	Decreased	PET film	Improve substrate binding by mutation to hydrophobic residue	(Ma, Yao, ... Yuan, 2018)
I179A	-	-	Decreased	PET bottle	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
I179A	-	-	Decreased	PET film	Verify role of residues that may be involved in catalytic mechanism by mutation.	(Han, Liu, ... Guo, 2017)
I179A	-	-	Decreased	PET film	Verify role of residues located in substrate binding site	(Joo, Cho, ... Kim, 2018)
I179Q	42.0	N.C.	Substantially Decreased	lcPET	Determine role of residues adjacent to active site	(Guo, Vanga, ... Syrén, 2022)
I179V	47.5	+1.5	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
I179V	-	-	Similar	PET film	Improve substrate binding by reducing steric hinderance	(Ma, Yao, ... Yuan, 2018)
I179V	-	-	Decreased	PET bottle	Structural bioinformatic hotspots for engineering	(Son, Joo, ... Kim, 2020)
I179F	-	-	Increased	PET film	Improve substrate binding by mutation to hydrophobic residue	(Ma, Yao, ... Yuan, 2018)
I179T	-	-	Substantially Decreased	PET bottle	Structural bioinformatic hotspots for engineering	(Son, Joo, ... Kim, 2020)
A180E	42.0	N.C.	Substantially Decreased	lcPET	Determine role of residues adjacent to active site	(Guo, Vanga, ... Syrén, 2022)
A180I	-	-	Similar	PET bottle	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
N183R	49.5	+3.5	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
S185H	55.5	+9.5	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
S185H	-	-	Increased	PET bottle	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
S185H	-	-	Decreased	PET film	Verify role of residues that may be involved in catalytic mechanism by mutation.	(Han, Liu, ... Guo, 2017)
R195Q	37.5	-11.2	-	-	MutCompute (Machine Learning)	(Lu, Diaz, ... Alper, 2022)
F200Y	54.4	+3	-	-	Premuse strategy	(Meng, Yang, ... Tian, 2021)

N204K	57.9	+9.2	-	-	MutCompute (Machine Learning)	(Lu, Diaz, ... Alper, 2022)
G205N	-	-	Decreased	PET bottle	Structural bioinformatic hotspots for engineering	(Son, Joo, ... Kim, 2020)
H208A (cHIS)	-	-	No Activity	-	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
H208A (cHIS)	-	-	No Activity	PET film	Verify role of residues located in substrate binding site	(Joo, Cho, ... Kim, 2018)
S209A	42.0	N.C.	Increased (pH 7.2, 30 °C)	lcPET	S209A allows W156 to access flat and perpendicular conformations for substrate binding	(Guo, Vanga, ... Syrén, 2022)
S209A	-	-	Decreased (pH 7.5, 30 °C)	PET film	Verify role of residues located in substrate binding site	(Joo, Cho, ... Kim, 2018)
S209F	47.5	+1.5	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
S209F	-	-	Decreased	PET film	Improve activity by mutation to hydrophilic residue	(Ma, Yao, ... Yuan, 2018)
S209F	-	-	Decreased	PET bottle	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
S209F	-	-	Decreased	PET film	Verify role of residues located in substrate binding site	(Joo, Cho, ... Kim, 2018)
S209V	-	-	Decreased	PET film	Improve substrate binding by mutation to hydrophobic residue	(Ma, Yao, ... Yuan, 2018)
S209T	-	-	Similar	PET bottle	Structural bioinformatic hotspots for engineering	(Son, Joo, ... Kim, 2020)
C210A	-	-	No Activity	pNPA	Determine role of active site disulfide by alanine substitution	(Fecker, Galaz-Davison, ... Ramírez-Sarmiento, 2018)
C210A	-	-	No Activity	PET bottle	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
C210S	-	-	No Activity	PET film	Verify role of residues that may be involved in catalytic mechanism by mutation.	(Han, Liu, ... Guo, 2017)
A211P	-	-	Decreased	PET film	Improve substrate release by reducing steric hinderance	(Ma, Yao, ... Yuan, 2018)
N212A	-	-	Decreased	PET bottle	Verify role of conserved residue by mutation.	(Liu, He, ... Bao, 2018)
N212A	-	-	Decreased	PET film	Verify role of residues located in substrate binding site	(Joo, Cho, ... Kim, 2018)

S213T	-	-	Increased	PET bottle	Structural bioinformatic hotspots for engineering	(Son, Joo, ... Kim, 2020)
N217D	-	-	Increased	PET bottle	Structural bioinformatic hotspots for engineering	(Son, Joo, ... Kim, 2020)
A219P	49.5	+3.5	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
T241Q	49	-2	-	-	Premuse strategy	(Meng, Yang, ... Tian, 2021)
T241Q	42.1	-0.4	-	-	PROSS strategy	(Rennison, Winther, and Varrone, 2021)
R251A	47.5	+1.5	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
R251A	-	-	Increased	PET film	Verify role of residues located in substrate binding site	(Joo, Cho, ... Kim, 2018)
C174A-C210A	-	-	No Activity	pNPA	Determine role of active site disulfide by alanine substitution	(Fecker, Galaz-Davison, ... Ramirez-Sarmiento, 2018)
C174A-C210A	-	-	Decreased	PET film	Verify role of residues located in substrate binding site	(Joo, Cho, ... Kim, 2018)
S185H I189F	-	-	Significantly Decreased	PET granules	Determine role of unique PETase residues	(Chen, Han, ... Guo, 2021)
W130H F200Y	61.2	+10.2	Increased	lcPET	Premuse strategy	(Meng, Yang, ... Tian, 2021)
S92D D157H	54.85	+6	Increased	PET film	Create H-bond found in Tf. Cut II (did not see in crystal structure)	(Son, Cho, ... Kim, 2019)
S92E D157H	56.02	+7.21	Increased	PET film	Create H-bond found in Tf. Cut II, longer sidechain than S92D, but did not observe H-bond in crystal structure. Improvement may be due to water-mediated H-bond.	(Son, Cho, ... Kim, 2019)
S92D P152A D157H	52.69	+3.88	Increased	PET film	Combination of stabilizing mutations	(Son, Cho, ... Kim, 2019)
S92E P152A D157H	53.56	+4.75	Increased	PET film	Combination of stabilizing mutations	(Son, Cho, ... Kim, 2019)
S92D D157H R251A	56.41	+7.6	Substantially Increased	PET film	Combination of stabilizing mutations	(Son, Cho, ... Kim, 2019)

	S92E D157H N217D	59	+10	Increased	PET bottle	Combining stabilizing single mutation with known mutation in Thermo PETase, but keeping R251 to help with salt bridge formation.	(Son, Joo, ... Kim, 2020)
	S92E D157H S213T N217D	60	+11	Increased	PET bottle	Combining stabilizing single mutations with known mutation in Thermo PETase, but keeping R251 to help with salt bridge formation.	(Son, Joo, ... Kim, 2020)
(Thermo PETase)	S92E D157H R251A	57.62	+8.81	Substantially Increased	PET film	Combination of stabilizing mutations	(Son, Cho, ... Kim, 2019)
	S92E N204K	57.0	+8.3	-	-	Combining MutCompute Stabilizing Mutations	(Lu, Diaz, ... Alper, 2022)
	T111D N204K	58.1	+9.4	-	-	Combining MutCompute Stabilizing Mutations	(Lu, Diaz, ... Alper, 2022)
	R195Q N204K	57.3	+8.6	-	-	Combining MutCompute Stabilizing Mutations	(Lu, Diaz, ... Alper, 2022)
	S92E R195Q	37.8	-10.9	-	-	Combining MutCompute Stabilizing Mutations	(Lu, Diaz, ... Alper, 2022)
	T111D R195Q	49.7	+1	-	-	Combining MutCompute Stabilizing Mutations	(Lu, Diaz, ... Alper, 2022)
	S92E T111D	49.1	+0.4	-	-	Combining MutCompute Stabilizing Mutations	(Lu, Diaz, ... Alper, 2022)
	S92E R195Q N204K	56.8	+8.1	Substantially Increased	lcPET	Combining MutCompute Stabilizing Mutations	(Lu, Diaz, ... Alper, 2022)
	T111D R195Q N204K	57.5	+8.8	-	-	Combining MutCompute Stabilizing Mutations	(Lu, Diaz, ... Alper, 2022)
	S92E T111D N204K	57.7	+9	-	-	Combining MutCompute Stabilizing Mutations	(Lu, Diaz, ... Alper, 2022)
	S92E T111D R195Q	48.0	-0.7	-	-	Combining MutCompute Stabilizing Mutations	(Lu, Diaz, ... Alper, 2022)
	S92E T111D R195Q N204K	56.9	+8.2	-	-	Combining MutCompute Stabilizing Mutations	(Lu, Diaz, ... Alper, 2022)

	S209F W130H	-	-	Slightly Increased	hcPET	Match Tf. Cut II Loop	(Austin, Allen, ... Beckham, 2018)
	Y58F W130H S209F	-	-	Similar	npPET	Site saturated mutagenesis of interaction site from docking/dynamics	(Pirillo, Orlando, ... Molla, 2021)
	Y58L W130H S209F	-	-	Similar	npPET	Site saturated mutagenesis of interaction site from docking/dynamics	(Pirillo, Orlando, ... Molla, 2021)
	W130H S209A	47.5	-3	Increased	npPET	Site saturated mutagenesis of interaction site from docking/dynamics	(Pirillo, Orlando, ... Molla, 2021)
	W130H S209E	-	-	Decreased	npPET	Site saturated mutagenesis of interaction site from docking/dynamics	(Pirillo, Orlando, ... Molla, 2021)
	W130H S209K	-	-	Decreased	npPET	Site saturated mutagenesis of interaction site from docking/dynamics	(Pirillo, Orlando, ... Molla, 2021)
	W130H S209F R251A	-	-	Increased	npPET	Site saturated mutagenesis of interaction site from docking/dynamics	(Pirillo, Orlando, ... Molla, 2021)
	W130H S209F R251V	-	-	Similar	npPET	Site saturated mutagenesis of interaction site from docking/dynamics	(Pirillo, Orlando, ... Molla, 2021)
	S92E W130H D157H S209A	55.3	+4.9	Increased	npPET mpPET	Combining S209A with known stabilizing mutations (S92E, D157H)	(Pirillo, Orlando, ... Molla, 2021)
	L88F Q90Y W130H G136A I139R A151I S159Q S185H R251A	75.0	+29	-	-	GRAPE strategy	(Cui, Chen, ... Wu, 2021)
(DuraPETase)	L88F Q90Y T111D W130H G136A I139R A151I	77.0	+31	Increased	hcPET	GRAPE strategy	(Cui, Chen, ... Wu, 2021)

	S159Q S185H R251A							
	R251A N204C-S253C	62.8	+14.7	-	-	Disulfide addition	(Zhong-Johnson, Voigt, and Sinskey, 2021)	
Is. PETase Variants	Thermo PETase + N204K	67.2	+18.5	-	-	Combining MutCompute stabilizing mutations with Thermo PETase	(Lu, Diaz, ... Alper, 2022)	
	Thermo PETase + T111D	60.8	+12.2	-	-	Combining MutCompute stabilizing mutations with Thermo PETase	(Lu, Diaz, ... Alper, 2022)	
	Thermo PETase + R195Q	59.8	+11.1	-	-	Combining MutCompute stabilizing mutations with Thermo PETase	(Lu, Diaz, ... Alper, 2022)	
	Thermo PETase + N204K T111D	66.7	+18	-	-	Combining MutCompute stabilizing mutations with Thermo PETase	(Lu, Diaz, ... Alper, 2022)	
(Fast PETase)	Thermo PETase + N204K R195Q	67.1	+18.4	Substantially Increased	lcPET	Combining MutCompute stabilizing mutations with Thermo PETase	(Lu, Diaz, ... Alper, 2022)	
	Thermo PETase + T111D R195Q	62.0	+13.3	-	-	Combining MutCompute stabilizing mutations with Thermo PETase	(Lu, Diaz, ... Alper, 2022)	
	Thermo PETase + T111D R195Q N204K	66.9	+18.2	-	-	Combining MutCompute stabilizing mutations with Thermo PETase	(Lu, Diaz, ... Alper, 2022)	
(TS PETase)	Thermo PETase + N204C-S253C	69.03	+20.22	Increased	PET film	Disulfide addition	(Zhong-Johnson, Voigt, and Sinskey, 2021)	
	Thermo PETase + K66N F172I	61.6	+16.6	Increased	Impranil	Error prone PCR + colonies incubated at 60 °C on Impranil agar plates	(Brott, Pfaff, ... Bornscheuer, 2022)	

Thermo PETase + S96N A197T	58.5	+13.4	Increased	Impranil	Error prone PCR + colonies incubated at 60 °C on Impranil agar plates	(Brott, Pfaff, ... Bornscheuer, 2022)
Thermo PETase + Q90L	58.6	+13.5	Increased	Impranil	Error prone PCR + colonies incubated at 60 °C on Impranil agar plates	(Brott, Pfaff, ... Bornscheuer, 2022)
Thermo PETase + T22A S96I S178I	58.5	+13.4	Increased	Impranil	Error prone PCR + colonies incubated at 60 °C on Impranil agar plates	(Brott, Pfaff, ... Bornscheuer, 2022)
Thermo PETase + K66N F172I N204C-S253C	70.9	+25.8	Decreased	npPET	Results from error prone PCR incorporated with Thermo PETase	(Brott, Pfaff, ... Bornscheuer, 2022)
Thermo PETase + N217D	54	+5	Increased (lower than Thermo PETase)	PET bottle	Combining stabilizing single mutation with Thermo PETase	(Son, Joo, ... Kim, 2020)
Dura PETase + N204C-S253C	81.1	+36.0	Decreased Increased	npPET lcPET	Disulfide addition combined with Dura PETase	(Brott, Pfaff, ... Bornscheuer, 2022; Liu, Liu, ... Wu, 2022)
Dura PETase + K66N S92E F172I R251A N204C-S253C	78.4	+33.3	Decreased	npPET	Results from error prone PCR incorporated with Dura PETase and disulfide bond	(Brott, Pfaff, ... Bornscheuer, 2022)
Dura PETase + K66N S92E F172I R251A	71.9	+26.8	Decreased	npPET	Results from error prone PCR incorporated with Dura PETase (lower T _m than Dura PETase)	(Brott, Pfaff, ... Bornscheuer, 2022)
Dura PETase + S92E	79.5	+30.8	-	-	Combining MutCompute stabilizing mutations with Thermo PETase	(Lu, Diaz, ... Alper, 2022)
Dura PETase + R195Q	78.3	+29.6	-	-	Combining MutCompute stabilizing mutations with Thermo PETase	(Lu, Diaz, ... Alper, 2022)

	Dura PETase + S92E N204K	83.4	+34.7	-	-	Combining MutCompute stabilizing mutations with Thermo PETase	(Lu, Diaz, ... Alper, 2022)
	Dura PETase + S92E R195Q	78.9	+30.2	-	-	Combining MutCompute stabilizing mutations with Thermo PETase	(Lu, Diaz, ... Alper, 2022)
	Dura PETase + R61D	-	-	Substantial Decrease	lcPET	Mutation of positive residue to negative residue to prove positive modification improves activity	(Liu, Liu, ... Wu, 2022)
	Dura PETase + H185S	-	-	Decreased	lcPET	Improve flexibility of key region	(Liu, Liu, ... Wu, 2022)
	Dura PETase + G214R	-	-	Similar	lcPET	Mutation of neutral region to positive region; based on docking results	(Liu, Liu, ... Wu, 2022)
	Dura PETase + S216R	-	-	Increased	lcPET	Mutation of neutral region to positive region; based on docking results	(Liu, Liu, ... Wu, 2022)
	Dura PETase + S216R N204C-N253C	-	-	Increased	lcPET	Combine known stabilizing/activity improving mutations	(Liu, Liu, ... Wu, 2022)
	Dura PETase + S216R H185S N204C-N253C	-	-	Substantially Increased	lcPET	Combine known stabilizing/activity improving mutations	(Liu, Liu, ... Wu, 2022)
	Dura PETase + H185S N204C-N253C	-	-	Increased	lcPET	Improve flexibility of key region; lower stability than Dura PETase + N204C-N253C, but higher activity	(Liu, Liu, ... Wu, 2022)
LCC (cSer = 130)	F208I	81.6	-3.1	Increased	pfPET	Site saturated mutagenesis of interaction site from docking	(Tournier, Topham, ... Marty, 2020)
	F208W	86.1	+1.4	Increased	pfPET	Site saturated mutagenesis of interaction site from docking	(Tournier, Topham, ... Marty, 2020)
	T61M	87.4	+2.7	Decreased	pfPET	Site saturated mutagenesis of interaction site from docking	(Tournier, Topham, ... Marty, 2020)
	Y92G	87.0	+2.3	Decreased	pfPET	Site saturated mutagenesis of interaction site from docking	(Tournier, Topham, ... Marty, 2020)
	N211D	87.9	+3.2	Similar	pfPET	Site saturated mutagenesis of interaction site from docking	(Tournier, Topham, ... Marty, 2020)
	N211M	88.0	+3.3	Similar	pfPET	Site saturated mutagenesis of interaction site from docking	(Tournier, Topham, ... Marty, 2020)
	H183S F187I	Reduced	-	Increased at lower temperatures	PET granules	Match unique PETase residues	(Chen, Han, ... Guo, 2021)

	D203C-S248C	94.5	+9.8	Decreased	pfPET	Disulfide Addition to replace Ca ²⁺ binding site	(Tournier, Topham, ... Marty, 2020)
	D203C-S248C F208I	90.9	+6.2	Increased	pfPET	Disulfide Addition + Known Active Mutation	(Tournier, Topham, ... Marty, 2020)
	D203C-S248C F208W	94.8	+10.1	Similar	pfPET	Disulfide Addition + Known Active Mutation	(Tournier, Topham, ... Marty, 2020)
	D203C-S248C F208I T61M	94.2	+9.5	Decreased	pfPET	Disulfide Addition + Active/Stability Increasing Mutations	(Tournier, Topham, ... Marty, 2020)
(ICCG)	D203C-S248C F208I Y92G	94.0	+9.3	Similar	pfPET	Disulfide Addition + Active/Stability Increasing Mutations	(Tournier, Topham, ... Marty, 2020)
	D203C-S248C F208I N211D	86.4	-0.1	Substantially Decreased	pfPET	Disulfide Addition + Active/Stability Increasing Mutations	(Tournier, Topham, ... Marty, 2020)
(ICCM)	D203C-S248C F208I N211M	94.5	+9.8	Increased	pfPET	Disulfide Addition + Active/Stability Increasing Mutations	(Tournier, Topham, ... Marty, 2020)
	D203C-S248C F208W T61M	98.0	+13.3	Decreased	pfPET	Disulfide Addition + Active/Stability Increasing Mutations	(Tournier, Topham, ... Marty, 2020)
	D203C-S248C F208W Y92G	98.0	+13.3	Similar	pfPET	Disulfide Addition + Active/Stability Increasing Mutations	(Tournier, Topham, ... Marty, 2020)
	D203C-S248C F208W N211D	88.6	+3.9	Substantially Decreased	pfPET	Disulfide Addition + Active/Stability Increasing Mutations	(Tournier, Topham, ... Marty, 2020)
(WCCM)	D203C-S248C F208W N211M	98.1	+13.4	Similar	pfPET	Disulfide Addition + Active/Stability Increasing Mutations	(Tournier, Topham, ... Marty, 2020)
	C240A-C257A	70.6	-15.6	-	-	Determine role of disulfide in LCC	(Sulaiman, You, ... Kanaya, 2014)
	+Glycosylation	-	+10	-	-	Reduce aggregation of LCC at high temperatures	(Shirke, White, ... Gross, 2018)
PET2 (cSer = 175)	R47C-G89C F105R E110K S156P G180A T296P	75.7	+8	Increased	lcPET	Addition of disulfide bond Surface charge modifications Backbone stabilizers	(Nakamura, Kobayashi, ... Iino, 2021)

	H228S F232I	-	-	Increased	PET granules	Match Is. PETase unique residues	(Chen, Han, ... Guo, 2021)
PET 6 (cSer = 163)	Y247A	-	-	Similar	Post- consumer PET	Remove steric hinderance of Tyr in substrate binding site	(Weigert, Perez- Garcia, ... Höcker, 2022)
	V91T S92A	-	-	Decreased	Post- consumer PET	Based on known stabilizing/activating mutations	(Weigert, Perez- Garcia, ... Höcker, 2022)
	T246S Y248N P259S S260N E251Q	-	-	Increased	Post- consumer PET	Match Is. PETase cHis loop	(Weigert, Perez- Garcia, ... Höcker, 2022)
PES-H1 (cSer = 130)	R47C-G89C	-	-	No Activity	PET film	Disulfide Addition	(Pfaff, Gao, ... Wei, 2022)
	F62A	77	N. C.	Substantially Decreased	PET film	Engineering of key residue involved with substrate binding	(Pfaff, Gao, ... Wei, 2022)
	F62Y	77	N. C.	Substantially Decreased	PET film	Engineering of key residue involved with substrate binding	(Pfaff, Gao, ... Wei, 2022)
	W155F	74	-3	Substantially Decreased	PET film	Engineering of key residue involved with substrate binding	(Pfaff, Gao, ... Wei, 2022)
	W155A	70	-7	Substantially Decreased	PET film	Engineering of key residue involved with substrate binding	(Pfaff, Gao, ... Wei, 2022)
	W155H	71	-6	Substantially Decreased	PET film	Engineering of key residue involved with substrate binding	(Pfaff, Gao, ... Wei, 2022)
	I178F	76	-1	Substantially Decreased	PET film	Engineering of key residue involved with substrate binding	(Pfaff, Gao, ... Wei, 2022)
	I178W	71	-6	Substantially Decreased	PET film	Engineering of key residue involved with substrate binding	(Pfaff, Gao, ... Wei, 2022)
	I178Y	77	N. C.	Substantially Decreased	PET film	Engineering of key residue involved with substrate binding	(Pfaff, Gao, ... Wei, 2022)
	L209K	56	-21	Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
	L209R	60	-17	Similar	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
	L209E	64	-13	Substantially Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
	L209P	73	-4	Substantially Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)

L209C	79	+2	Substantially Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
L209I	75	-2	Substantially Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
L209T	79	+2	Substantially Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
L209H	80	+3	Substantially Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
L209D	77	N. C.	Substantially Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
L209F	81	+4	Substantially Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
L209Y	81	+4	Substantially Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
L209V	77	N. C.	Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
L209M	78	+1	Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
L209W	81	+4	Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
L209S	79	+2	Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
L209Q	79	+2	Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
L209N	79	+2	Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
L209A	77	N. C.	Decreased	PET film	Selective saturation mutagenesis on a known engineering hotspot	(Pfaff, Gao, ... Wei, 2022)
R204C-S250C	76.8	-0.3	Similar	PET film	Disulfide Addition	(Pfaff, Gao, ... Wei, 2022)
L92F Q94Y	78.2	+1.1	Increased	PET film	Based on mutations in Dura PETase	(Pfaff, Gao, ... Wei, 2022)
L92F Q94Y R204C-S250C	81.1	4.8	Decreased	PET film	Combining known stabilizing/activating mutations	(Pfaff, Gao, ... Wei, 2022)
Bb PETase (cSer = 134)	54	+3	Similar	mpPET	Compare full length to truncated Bb PETase	(Sagong, Kim, ... Kim, 2022)
+ N Terminal L49T	51.5	+0.5	Decreased	mpPET	Match Is. PETase residue	(Sagong, Kim, ... Kim, 2022)

	+ N Terminal S51R	52.5	+1.5	Decreased	mpPET	Match Is. PETase residue	(Sagong, Kim, ... Kim, 2022)
	+N Terminal S166N	55	+4	Increased	mpPET	Match Is. PETase residue	(Sagong, Kim, ... Kim, 2022)
	+N Terminal T169I	56	+5	Increased	mpPET	Match Is. PETase residue	(Sagong, Kim, ... Kim, 2022)
	+N Terminal M194I	58	+7	Increased	mpPET	Match Is. PETase residue	(Sagong, Kim, ... Kim, 2022)
	+N Terminal N196G	55	+3	Increased	mpPET	Match Is. PETase residue	(Sagong, Kim, ... Kim, 2022)
(NIIG)	+N Terminal S166N T169I M194I N196G	63	+12	Increased	mpPET	Combine stabilizing mutations	(Sagong, Kim, ... Kim, 2022)
Cbotu_Est A (cSer = 181)	S126A	Decreased	-	Decreased	PET film	Engineering of zinc binding domain	(Biundo, Reich, ... Guebitz, 2018)
	W128A	Decreased	-	Decreased	PET film	Engineering of zinc binding domain	(Biundo, Reich, ... Guebitz, 2018)
	F153Y	Increased	-	Decreased	PET film	Engineering of zinc binding domain	(Biundo, Reich, ... Guebitz, 2018)
	S198A	Increased	-	Decreased	PET film	Engineering of zinc binding domain	(Biundo, Reich, ... Guebitz, 2018)
	W273H	Increased	-	Decreased	PET film	Engineering of zinc binding domain	(Biundo, Reich, ... Guebitz, 2018)
	del-70	Decreased	-	Increased	PET film	Removal of N-terminal domain	(Biundo, Reich, ... Guebitz, 2018)
	del-70 S126A	Decreased	-	Increased (Lower than del-70)	PET film	Engineering of zinc binding domain	(Biundo, Reich, ... Guebitz, 2018)
	del-70 W128A	Decreased	-	Increased	PET film	Engineering of zinc binding domain	(Biundo, Reich, ... Guebitz, 2018)
	del-70 F153Y	Decreased	-	Increased (Lower than del-70)	PET film	Engineering of zinc binding domain	(Biundo, Reich, ... Guebitz, 2018)
	del-70 S198A	Decreased	-	Increased	PET film	Engineering of zinc binding domain	(Biundo, Reich, ... Guebitz, 2018)
	del-70 W273A	Decreased	-	Increased	PET film	Engineering of zinc binding domain	(Biundo, Reich, ... Guebitz, 2018)

FsC (cSer = 104)	L166A	-	-	Substantially Increased	PET fabric	Binding free energy changes following mutation to a residue in active site (MD)	(Araújo, Silva, ... Cavaco-Paulo, 2007)
	L173A	-	-	Decreased	PET fabric	Binding free energy changes following mutation to a residue in active site (MD)	(Araújo, Silva, ... Cavaco-Paulo, 2007)
	V168A	-	-	Increased	PET fabric	Binding free energy changes following mutation to a residue in active site (MD)	(Araújo, Silva, ... Cavaco-Paulo, 2007)
	L65A	-	-	Increased	PET fabric	Binding free energy changes following mutation to a residue in active site (MD)	(Araújo, Silva, ... Cavaco-Paulo, 2007)
	N68A	-	-	Increased	PET fabric	Binding free energy changes following mutation to a residue in active site (MD)	(Araújo, Silva, ... Cavaco-Paulo, 2007)
Bioinformatics Based	Global Changes	-	-	-	-	Ancestral Sequence Reconstruction	(Perlmutter and Theobald, 2020)
Fusion Proteins	Is. PETase + CBM66	-	-	-	-	Improved protein expression in <i>E. coli</i>	(Ko, Kang, ... Sung, 2021)
	Is. PETase + Next Tag	-	-	-	-	Improved protein expression in <i>E. coli</i>	(Jo, 2021)
	Is. PETase + Trx domain	-	-	Slightly Increased	pwPET	Improved protein production in <i>E. coli</i>	(Wagner-Egea, Tosi, ... Linares-Pastén, 2021)
	Is. PETase + MHETase	-	-	Increased	lcPET	Addition of MHETase to Is. PETase, which is capable of breaking down MHET	(Knott, Erickson, ... McGeehan, 2020)
	Is. PETase + engineered pelB	-	-	-	-	Improved protein expression in <i>E. coli</i>	(Shi, Liu, ... Zhu, 2021)
	Is. PETase + YeeJ translocator domain	-	-	Increased compared to free PETase	hcPET	Surface display of PETase on <i>E. coli</i>	(Gercke, Furtmann, ... Jose, 2021)
	Is. PETase + MalE	-	-	-	-	Attempt for extracellular production of Is. PETase in <i>E. coli</i> . Resulted in very little excretion.	(Seo, Kim, ... Kim, 2019)
	Is. PETase + LamB	-	-	-	-	Attempt for extracellular production of Is. PETase in <i>E. coli</i> . Resulted in little excretion.	(Seo, Kim, ... Kim, 2019)
	Is. PETase + SurA	-	-	-	-	Attempt for extracellular production of Is. PETase in <i>E. coli</i> . Resulted in no excretion.	(Seo, Kim, ... Kim, 2019)
Is. PETase + DsbA	-	-	-	-	Attempt for extracellular production of Is. PETase in <i>E. coli</i> . Resulted in no excretion.	(Seo, Kim, ... Kim, 2019)	

Is. PETase + TolB	-	-	-	-	Attempt for extracellular production of Is. PETase in <i>E. coli</i> . Resulted in no excretion.	(Seo, Kim, ... Kim, 2019)
Is. PETase + pelB	-	-	-	-	Attempt for extracellular production Is. PETase in <i>E. coli</i> . Resulted in no excretion.	(Seo, Kim, ... Kim, 2019)
Is. PETase + <i>Bacillus subtilis</i> signal peptide	-	-	-	-	Extracellular production of Is. PETase in <i>Bacillus subtilis</i> using a native signal peptide	(Huang, Cao, ... Liu, 2018)

B. Cut-PETase Multiple Sequence Alignment

N	P	Y	E	R	G	P	A	P	T	E	A	S	I	E	A	S	R	G	P	F	A	V	S	E	T	S	V	S	S	L	X	V	S	G	F	G	G	G	T	I	Y	Y	P	T	S	T	S	E	G	
N	P	Y	E	R	G	P	N	P	T	D	A	L	L	E	A	R	S	G	P	F	S	V	S	E	N	V	S	R	L	S	A	S	G	F	G	G	G	T	I	Y	Y	P	R	E	N	N	-	-	G	
N	P	Y	A	R	G	P	N	P	T	A	A	S	L	E	A	S	A	G	P	F	T	V	R	S	F	T	V	S	R	-	-	P	S	G	Y	G	A	G	T	V	Y	Y	P	T	N	A	G	-	-	G
N	P	Y	Q	R	G	P	N	P	T	R	S	A	L	T	A	-	D	G	P	F	S	V	A	T	Y	T	V	S	R	L	S	V	S	G	F	G	G	G	V	I	Y	Y	P	T	G	T	S	L	-	-
N	P	Y	E	R	G	P	D	P	T	E	D	S	I	E	A	I	R	G	P	F	S	V	A	T	E	R	V	S	S	F	-	A	S	G	F	G	G	G	T	I	Y	Y	P	R	E	T	D	E	G	
N	P	Y	E	R	G	P	N	P	T	D	A	L	L	E	A	R	S	G	P	F	S	V	S	E	E	R	A	S	R	F	G	A	D	G	F	G	G	G	T	I	Y	Y	P	R	E	N	N	-	-	
N	P	Y	E	R	G	P	N	P	T	E	S	M	L	E	A	R	S	G	P	F	S	V	S	E	E	R	A	S	R	F	G	A	D	G	F	G	G	G	T	I	Y	Y	P	R	E	N	N	-	-	
N	P	Y	E	R	G	P	A	P	T	N	A	S	I	E	A	S	R	G	P	Y	A	T	S	Q	T	S	V	S	S	L	V	A	S	G	F	G	G	G	T	I	Y	Y	P	T	S	T	A	D	G	
N	P	Y	Q	R	G	P	D	P	T	E	S	L	L	R	A	A	R	G	P	F	A	V	S	E	Q	S	V	S	R	L	S	V	S	G	F	G	G	G	R	I	Y	Y	P	T	T	S	Q	G		
N	P	Y	E	R	G	P	N	P	T	D	A	L	L	E	A	R	S	G	P	F	S	V	S	E	E	R	A	S	R	F	G	A	D	G	F	G	G	G	T	I	Y	Y	P	R	E	N	N	-	-	
N	P	Y	E	R	G	P	N	P	T	N	S	I	E	A	L	R	G	P	F	R	V	D	E	E	R	V	S	R	L	Q	A	R	G	F	G	G	G	T	I	Y	Y	P	T	D	N	N	-	-		
N	P	Y	E	R	G	P	N	P	T	E	S	M	L	E	A	R	S	G	P	F	S	V	S	E	E	R	A	S	R	F	G	A	D	G	F	G	G	G	T	I	Y	Y	P	R	E	N	N	-	-	
N	P	Y	E	R	G	P	N	P	T	E	S	M	L	E	A	R	S	G	P	F	S	V	S	E	E	R	A	S	R	L	G	A	D	G	F	G	G	G	T	I	Y	Y	P	R	E	N	N	-	-	
-	-	F	E	R	G	P	A	P	T	E	A	S	I	E	A	A	K	G	P	F	A	I	D	R	I	E	V	P	A	G	S	G	T	G	F	N	S	G	T	I	Y	Y	P	T	S	T	A	E	G	
-	-	Y	Q	R	G	P	D	P	T	S	V	S	F	L	E	A	D	R	G	Q	Y	S	V	R	S	S	R	V	S	S	L	-	V	S	G	F	G	G	G	T	I	Y	Y	P	T	G	T	-	G	
N	P	Y	E	R	G	P	D	P	T	D	S	S	V	E	A	L	R	G	P	Y	S	V	D	D	T	S	V	S	S	L	-	V	S	G	F	G	G	G	T	V	Y	Y	P	T	G	T	N	E	-	
N	P	Y	E	R	G	P	A	P	T	D	A	V	L	E	A	S	R	G	P	F	A	T	T	S	Y	S	V	S	G	L	S	V	T	G	F	G	G	G	V	V	Y	Y	P	T	S	T	S	E	G	
N	P	Y	E	R	G	P	A	P	T	V	A	L	L	E	A	S	R	G	P	F	T	T	A	S	Q	N	V	S	S	L	S	V	T	G	F	G	G	G	V	I	Y	Y	P	T	D	T	S	Q	G	
N	P	Y	E	R	G	P	D	P	T	V	Q	S	I	E	A	V	R	G	P	F	S	V	S	E	T	N	V	S	R	L	V	T	G	F	G	G	G	T	I	Y	Y	P	R	D	T	S	Q	G		
N	P	Y	E	R	G	P	A	P	T	N	A	V	L	E	A	S	R	G	P	F	A	T	S	T	N	V	S	S	L	-	V	S	G	F	G	G	G	V	I	Y	Y	P	T	T	T	S	E	G		
N	P	Y	E	R	G	P	D	P	T	V	A	S	I	E	A	R	R	G	P	F	A	V	A	E	D	S	V	S	R	Y	V	V	T	G	F	G	G	G	T	V	Y	Y	P	T	S	T	A	E	G	
N	P	Y	E	R	G	P	A	P	T	E	S	S	V	T	A	V	R	G	Y	F	D	T	D	T	D	T	V	S	S	L	-	V	S	G	F	G	G	G	T	I	Y	Y	P	T	D	T	S	E	G	
N	P	Y	E	R	G	P	A	P	T	T	A	S	I	E	A	A	R	G	P	F	A	I	S	Q	T	S	V	S	S	L	-	V	G	G	F	G	G	G	T	I	Y	Y	P	T	S	T	A	S	G	
N	P	Y	E	R	G	P	N	P	T	E	S	S	I	E	A	T	R	G	Y	Y	S	V	S	D	E	S	V	S	S	L	-	A	P	G	F	G	G	G	T	I	Y	Y	P	T	T	R	T	D	G	
N	P	Y	E	R	G	P	A	P	T	V	A	S	I	E	A	T	R	G	P	F	A	I	A	Q	T	T	V	S	A	S	S	V	S	G	F	R	G	G	G	T	I	Y	Y	P	T	S	T	A	E	G
N	P	H	E	R	G	P	A	P	T	E	Q	S	V	T	A	P	R	G	H	F	D	T	A	T	A	S	V	S	S	L	-	V	S	G	F	G	G	G	T	I	Y	Y	P	T	D	T	S	E	G	
G	P	Y	W	R	G	P	A	P	T	E	A	S	V	T	A	T	T	G	P	F	R	V	T	E	A	R	I	T	G	-	-	A	R	G	F	G	G	G	R	I	Y	Y	P	T	D	T	S	Q	G	
N	P	Y	Q	R	G	P	E	P	T	E	Q	S	V	T	A	E	R	G	P	F	A	T	S	T	V	G	V	P	A	G	S	G	T	G	F	N	R	G	T	I	Y	Y	P	T	D	T	S	Q	G	
D	T	Y	Q	R	G	P	D	P	T	A	S	S	V	L	S	-	H	G	T	F	A	L	S	T	T	S	V	S	S	A	-	V	S	G	F	G	G	G	T	I	Y	Y	P	S	S	T	D	E	G	
S	A	T	E	R	G	L	A	P	T	E	A	N	I	T	G	-	D	G	S	Y	G	V	S	A	T	I	T	G	-	-	A	S	G	F	G	G	G	V	Y	Y	P	N	A	T	E	-	-			
N	P	Y	Q	R	G	P	N	P	T	E	A	S	I	T	A	A	R	G	P	F	N	T	A	E	I	T	V	S	R	L	S	V	T	G	F	G	G	G	K	I	Y	Y	P	T	T	T	S	E	G	
N	P	Y	E	R	G	P	A	P	T	E	A	S	I	E	A	A	R	G	P	F	A	T	S	Q	F	S	V	S	S	L	S	V	T	G	F	G	G	G	T	V	Y	Y	P	T	S	T	A	E	G	
N	P	Y	Q	R	G	P	D	P	T	D	A	L	L	E	A	P	R	G	P	F	A	T	A	S	T	T	V	S	R	L	S	V	S	G	F	G	G	G	V	I	Y	Y	P	T	D	T	S	Q	G	
N	P	Y	E	R	G	P	N	P	T	E	S	L	I	E	A	A	R	G	P	F	A	T	S	Q	T	T	V	S	S	L	S	V	T	G	F	G	G	G	V	I	Y	Y	P	T	D	T	S	Q	G	
N	P	Y	E	R	G	P	D	P	T	E	Q	L	I	E	S	A	R	G	P	F	A	T	A	Q	T	R	V	S	S	L	S	V	T	G	F	G	G	G	V	I	Y	Y	P	T	D	T	S	Q	G	
S	P	Y	E	R	G	P	A	P	T	E	Q	S	V	S	A	K	L	G	P	F	A	F	T	Q	V	P	A	P	A	N	A	S	A	G	F	N	O	G	T	V	Y	Y	P	N	D	T	G	Q	G	
-	-	Y	Q	R	G	P	A	P	T	E	A	S	V	R	A	D	R	G	P	F	Q	V	A	R	V	N	V	P	G	-	-	A	G	F	G	S	G	T	I	Y	Y	P	T	D	T	S	Q	G		
N	P	Y	E	R	G	P	D	P	T	E	R	S	V	T	A	A	R	G	P	F	A	I	A	Q	M	E	V	S	G	G	V	G	A	G	F	N	S	G	T	V	F	Y	P	T	D	T	G	Q	G	
-	-	Y	E	R	G	P	A	P	T	W	A	G	I	R	T	P	L	G	P	F	S	Y	S	K	V	T	V	T	N	A	E	A	N	G	F	G	G	G	T	I	Y	Y	P	N	D	T	T	Q	G	
N	P	Y	Q	R	G	P	D	P	T	E	R	S	V	T	A	A	R	G	P	F	A	I	D	E	I	S	V	N	G	G	I	G	A	G	F	N	R	G	T	I	F	Y	P	T	D	R	S	Q	G	
N	P	F	E	R	G	P	A	P	T	E	S	S	I	E	A	L	R	G	P	Y	A	V	S	E	T	V	S	R	L	A	V	S	G	F	G	G	G	T	I	Y	Y	P	T	S	T	A	D	G		
N	P	Y	E	R	G	P	D	P	T	E	D	S	I	E	A	D	R	R	G	P	Y	S	V	D	E	N	V	S	S	L	-	V	P	G	F	G	G	G	T	I	Y	Y	P	S	R	T	E	-		
N	P	Y	E	R	G	P	A	P	T	V	S	S	I	E	A	L	R	G	P	F	A	V	S	E	T	S	V	S	S	L	-	V	G	G	F	G	G	G	T	I	Y	Y	P	T	S	T	T	S	G	
N	P	Y	E	R	G	P	A	P	T	E	S	S	I	E	A	L	R	G	P	Y	S	V	A	E	T	E	V	S	S	L	A	V	T	G	F	G	G	G	T	I	Y	Y	P	T	S	T	S	D	G	
N	P	Y	E	R	G	P	A	P	T	E	A	G	I	E	A	L	R	G	P	Y	A	V	S	E	T	S	V	S	S	L	A	A	S	G	F	G	G	G	T	I	Y	Y	P	T	S	T	A	D	G	
N	P	Y	E	R	G	P	D	P	T	S	V	A	I	L	E	A	S	R	G	P	F	S	V	S	T	Q	N	V	S	S	L	S	V	T	G	F	G	G	G	V	I	Y	Y	P	T	T	A	D	G	
N	P	Y	E	R	G	P	A	P	T	L	O	A	L	Q	A	S	R	G	P	Y	A	V	S	T	T	S	V	S	R	L	S	V	S	G	F	G	G	G	T	I	Y	Y	P	T	T	T	A	D	G	
N	P	Y	Q	R	G	P	N	P	T	Q	S	T	L	D	A	S	R	G	P	Y	A	V	A	T	E	N	V	S</																						

	T	F	G	A	V	A	I	S	P	G	Y	T	A	T	Q	S	S	I	A	W	L	G	P	R	L	A	S	Q	G	F	V	V	I	T	I	D	T	N	T	T	L	D	Q	P	D	S	R	G	R	Q
	T	Y	G	A	V	A	I	S	P	G	Y	T	G	T	E	A	S	I	A	W	L	G	E	R	I	A	S	H	G	F	V	V	I	T	I	D	T	I	T	T	L	D	Q	P	D	S	R	A	E	Q
60	T	V	G	A	I	A	I	V	P	G	Y	T	A	R	Q	S	S	I	K	W	W	G	P	R	L	A	S	H	G	F	V	V	I	T	I	D	T	N	S	T	L	D	Q	P	S	S	R	S	S	Q
	T	F	G	G	I	A	M	S	P	G	Y	T	A	D	A	S	S	L	A	W	L	G	R	R	L	A	S	H	G	F	V	V	L	V	I	N	T	N	S	R	F	D	Y	P	D	S	R	A	S	Q
	T	F	G	A	V	A	V	A	P	G	F	T	A	S	Q	G	S	M	S	W	Y	G	E	R	V	A	S	Q	G	F	I	V	F	T	I	D	T	N	T	R	L	D	Q	P	G	Q	R	G	R	Q
	T	Y	G	A	V	A	I	S	P	G	Y	T	G	T	Q	A	S	V	A	W	L	G	K	R	I	A	S	H	G	F	V	V	I	T	I	D	T	N	T	T	L	D	Q	P	D	S	R	A	R	Q
	T	Y	G	A	I	A	I	S	P	G	Y	T	G	T	Q	S	S	I	A	W	L	G	E	R	I	A	S	H	G	F	V	V	I	A	I	D	T	N	T	T	L	D	Q	P	D	S	R	A	R	Q
	T	F	G	A	V	I	S	P	G	F	T	A	Y	Q	S	S	I	A	W	L	G	P	R	L	A	S	H	G	F	V	V	F	T	I	D	T	N	T	T	L	D	Q	P	D	S	R	A	R	Q	
	T	F	G	A	I	A	I	S	P	G	F	T	A	S	W	S	S	L	A	W	L	G	P	R	L	A	S	H	G	F	V	V	I	G	I	E	T	N	T	R	L	D	Q	P	D	S	R	G	R	Q
	T	Y	G	A	V	A	I	S	P	G	Y	T	G	T	Q	A	S	V	A	W	L	G	E	R	I	A	S	H	G	F	V	V	I	T	I	D	T	N	T	T	L	D	Q	P	D	S	R	A	R	Q
	T	F	G	A	V	A	I	S	P	G	Y	T	G	T	Q	S	S	I	S	W	L	G	E	R	L	A	S	H	G	F	V	V	M	T	I	D	T	N	T	T	L	D	Q	P	D	S	R	A	S	Q
	T	Y	G	A	I	A	I	S	P	G	Y	T	G	T	Q	S	S	I	A	W	L	G	E	R	I	A	S	H	G	F	V	V	I	A	I	D	T	N	T	T	L	D	Q	P	D	S	R	A	R	Q
	T	Y	G	A	I	A	I	S	P	G	Y	T	G	T	Q	S	S	I	A	W	L	G	E	R	I	A	S	H	G	F	V	V	I	A	I	D	T	N	T	T	L	D	Q	P	D	S	R	A	R	Q
	T	F	G	A	V	A	I	S	P	G	F	V	S	P	K	S	W	I	D	W	Y	G	P	R	L	A	S	Q	G	F	V	V	M	T	L	E	T	F	S	Y	F	D	A	P	D	G	R	A	D	Q
	T	M	G	A	V	V	V	I	P	G	F	V	S	A	E	S	S	I	D	W	W	G	P	K	L	A	S	Y	G	F	V	V	M	T	I	D	T	N	T	G	F	D	Q	P	P	S	R	A	R	Q
	T	F	G	A	V	V	I	A	P	G	Y	T	A	S	S	S	S	M	S	W	M	G	E	R	L	A	S	Q	G	F	V	I	F	T	I	D	T	N	T	R	Y	D	Q	P	D	S	R	G	D	Q
	T	F	G	A	I	A	I	S	P	G	Y	T	A	T	W	S	S	L	S	W	L	G	P	R	I	A	S	H	G	F	V	V	I	G	I	E	T	N	T	L	D	Q	P	D	S	R	G	R	Q	
	T	F	G	A	I	A	I	S	P	G	Y	T	A	A	W	S	S	I	S	W	L	G	P	R	I	A	S	H	G	F	V	V	I	G	I	E	T	N	S	R	L	D	Q	P	D	S	R	G	R	Q
	T	F	G	A	V	A	I	A	P	G	F	T	A	S	E	G	S	M	A	W	Y	G	P	R	I	A	S	Q	G	F	V	V	F	T	I	D	T	L	T	R	L	D	Q	P	D	S	R	G	R	Q
	T	F	G	A	V	A	I	S	P	G	Y	T	A	S	W	S	S	M	D	W	L	G	P	R	I	A	S	H	G	F	V	V	I	G	I	E	T	N	S	R	L	D	Q	P	A	S	R	G	R	Q
	T	F	G	A	V	A	V	A	P	G	Y	T	A	T	E	S	S	M	A	W	L	G	P	R	L	A	S	Q	G	F	V	V	F	T	I	N	T	R	T	T	S	D	Q	P	S	S	R	G	D	Q
	T	F	G	G	V	V	I	A	P	G	Y	T	A	S	Q	S	S	M	A	W	M	G	H	R	I	A	S	Q	G	F	V	V	F	T	I	D	T	I	T	R	Y	D	Q	P	D	S	R	G	R	Q
40	T	F	G	A	V	A	I	S	P	G	F	T	G	T	Q	A	T	I	A	W	L	G	P	R	I	A	S	Q	G	F	V	V	F	T	I	D	T	T	T	V	F	D	Q	P	D	S	R	A	T	Q
	T	F	G	A	V	A	V	V	P	G	Y	T	A	S	E	S	T	I	E	W	L	G	A	R	V	A	S	Q	G	F	V	V	F	T	I	D	T	D	T	R	L	D	Q	P	R	Q	R	G	D	Q
	T	F	G	A	V	A	I	S	P	G	F	T	A	S	Q	S	S	V	A	W	L	G	P	R	L	A	S	Q	G	F	V	V	I	T	I	D	T	L	S	V	Y	D	Q	P	A	S	R	G	T	Q
	T	F	G	G	V	V	V	A	P	G	Y	T	A	S	E	S	S	M	S	W	Y	G	H	R	F	A	S	Q	G	F	V	V	F	T	I	D	T	N	T	R	Y	D	Q	P	N	S	R	G	R	Q
	T	F	G	A	V	A	I	S	P	G	F	T	A	S	Y	W	S	S	L	D	W	L	G	P	R	L	A	S	Q	G	F	V	V	I	G	I	E	T	N	T	T	L	D	Q	P	A	Q	R	D	Q
	T	F	G	A	I	A	V	I	P	G	F	T	G	P	E	S	S	I	A	W	F	G	P	H	L	A	S	H	G	F	V	V	M	T	L	T	T	S	S	G	W	D	F	P	A	S	R	T	N	Q
	T	Y	G	G	V	V	L	A	P	G	F	T	A	S	S	W	M	Y	D	S	I	A	R	R	A	A	S	H	G	F	V	V	F	A	I	D	T	N	S	I	Y	D	Q	P	G	S	R	G	T	Q
	R	F	P	V	V	A	I	S	P	G	Y	T	E	R	W	S	S	F	A	W	L	G	R	R	L	A	S	W	G	F	V	V	V	G	I	E	T	N	S	L	F	D	Q	P	N	S	R	G	T	Q
	T	F	G	A	I	A	I	S	P	G	F	T	A	Y	W	S	S	L	E	W	L	G	H	R	L	A	S	Q	G	F	V	V	I	G	I	E	T	N	T	T	L	D	Q	P	D	Q	R	G	Q	Q
	T	F	G	A	V	A	I	S	P	G	Y	T	A	R	W	S	S	L	D	W	L	G	P	R	I	A	S	Q	G	F	V	V	I	G	I	E	T	N	T	L	Y	D	Q	P	S	S	R	G	R	Q
	T	F	G	G	I	A	I	S	P	G	Y	T	A	R	W	S	S	L	D	W	L	G	P	R	I	A	S	H	G	F	V	V	I	G	I	E	T	N	S	R	Y	D	Q	P	A	S	R	G	R	Q
	T	F	G	G	I	A	I	S	P	G	Y	T	A	A	W	S	S	L	D	W	L	G	P	R	L	A	S	H	G	F	V	V	I	G	I	E	T	N	T	R	L	D	Q	P	A	S	R	G	D	Q
	T	F	G	A	V	A	I	S	P	G	F	T	A	D	W	T	S	L	D	W	L	G	P	R	L	A	S	H	G	F	V	V	I	G	I	D	T	I	T	R	L	D	Q	P	D	S	R	G	R	Q
	T	F	G	A	I	A	I	T	P	G	F	I	S	P	E	S	S	I	A	W	L	G	P	R	L	A	S	N	G	F	V	V	I	T	L	T	N	T	G	F	D	Q	P	N	G	R	A	D	Q	
	T	F	G	G	V	A	V	S	P	G	F	L	S	P	E	I	L	I	A	W	L	G	P	R	L	A	S	H	G	F	V	V	I	T	F	T	T	N	T	P	T	D	Q	P	D	S	R	G	S	Q
	T	F	G	A	I	A	A	I	P	G	F	V	S	P	E	A	I	V	R	W	F	G	P	R	L	A	S	Q	G	F	V	V	I	T	L	T	S	D	G	A	T	D	T	P	A	S	R	G	D	Q
	T	F	G	G	V	A	I	S	P	G	Y	T	F	G	P	E	G	H	V	A	W	L	G	P	R	L	A	S	Q	G	F	I	V	F	T	I	A	T	N	S	V	Y	D	Q	P	T	E	R	G	Q
	T	F	G	A	V	A	V	I	P	G	F	L	S	P	E	S	L	V	R	W	F	G	P	R	L	A	S	Q	G	F	V	V	M	T	L	T	N	G	L	T	D	T	P	E	S	R	S	E	Q	
	T	F	G	A	V	A	V	S	P	G	Y	T	G	T	Q	S	S	I	A	W	L	G	P	R	L	A	S	Q	G	F	V	V	F	T	I	D	T	L	T	T	L	D	Q	P	D	S	R	G	N	Q
20	T	F	G	G	V	A	I	A	P	G	Y	T	A	G	E	S	S	M	S	W	Y	G	P	R	L	A	S	Q	G	F	V	V	F	T	I	D	T	N	T	R	F	D	Q	P	G	S	R	G	D	Q
	T	F	G	A	V	A	V	S	P	G	Y	T	G	T	Q	S	S	I	S	W	L	G	P	R	L	A	S	Q	G	F	V	V	F	T	I	D	T	N	T	I	Y	D	Q	P	D	S	R	A	S	Q
	T	F	G	A	V	V	I	S	P	G	F	T	A	L	Q	S	S	I	A	W	L	G	P	R	L	A	S	Q	G	F	V	V	F	T	I	D	T	N	T	T	V	D	Q	P	D	S	R	G	R	Q
	T	F	G	A	I	A	I	S	P	G	Y	T	A	R	E	S	S	I	A	W	L	G	P	R	L	A	S	Q	G	F	V	V	F	T	I	D	T	N	S	V	Y	D	Q	P	A	S	R	G	R	Q
	T	F	G	A	I	A	I	S	P	G	Y	T	A	A	W	S	S	I	D	W	L	G	P	R	L	A	S	H	G	F	V	V	I	G	I	E	T	I	T	R	L	D	Q	P	D	S	R	G	R	Q
	T	F	G	A	I	A	I	S	P	G	Y	T	A	A	W	S	S	I	S	W	L	G	P	R	L	A	S	H	G	F	V	V	I	G	I	E	T	L												

/s. PETase Loop Insert #1

Catalytic Serine

L	L	A	A	L	D	Y	L	T	Q	R	-	-	-	S	S	V	R	S	R	I	D	A	S	R	L	A	V	M	G	H	S	M	G	G	G	G	S	L	E	A	A	X	S	R	P	S	L				
L	N	A	A	L	N	H	M	I	N	R	-	-	-	A	S	S	T	V	R	S	R	I	D	S	S	R	L	A	V	M	G	H	S	M	G	G	G	G	S	L	R	L	A	S	Q	R	P	D	L		
Q	M	A	A	L	R	Q	V	A	S	L	N	G	T	S	S	S	P	I	Y	G	K	V	D	T	A	R	M	L	A	V	M	G	H	S	M	G	G	G	G	S	L	L	I	S	A	A	N	N	P	S	L
L	S	A	A	L	N	Y	L	R	T	S	-	-	-	S	P	S	A	V	R	A	R	L	D	A	N	R	L	A	V	A	G	H	S	M	G	G	G	G	T	L	R	I	A	E	Q	N	P	S	L		
L	L	A	A	L	D	Y	L	V	E	R	-	-	-	S	D	R	K	V	R	E	R	L	D	P	N	R	L	A	V	M	G	H	S	M	G	G	G	G	S	L	E	A	T	V	M	R	P	S	L		
L	N	A	A	L	D	Y	M	I	N	D	-	-	-	A	S	S	A	V	R	S	R	I	D	S	S	R	L	A	V	M	G	H	S	M	G	G	G	G	S	L	R	L	A	S	Q	R	P	D	L		
L	N	A	A	L	D	Y	M	L	T	D	-	-	-	A	S	S	A	V	R	N	R	I	D	A	S	R	L	A	V	M	G	H	S	M	G	G	G	G	T	L	R	L	A	S	Q	R	P	D	L		
L	L	S	A	L	D	Y	L	T	Q	R	-	-	-	-	-	S	S	V	R	T	R	V	D	A	T	R	L	G	V	M	G	H	S	M	G	G	G	G	S	L	E	A	A	K	S	R	T	S	L		
L	L	A	A	L	D	Y	L	T	Q	R	-	-	-	-	-	S	S	V	R	N	R	I	D	A	S	R	L	A	V	A	G	H	S	M	G	G	G	G	T	L	E	A	A	K	S	R	T	S	L		
L	N	A	A	L	D	Y	M	I	N	D	-	-	-	A	S	S	A	V	R	S	R	I	D	S	S	R	L	A	V	M	G	H	S	M	G	G	G	G	T	L	R	L	A	S	Q	R	P	D	L		
L	D	A	A	L	D	Y	M	V	E	D	-	-	-	S	S	Y	S	V	R	N	R	I	D	S	S	R	L	A	A	M	G	H	S	M	G	G	G	G	T	L	R	L	A	E	R	R	P	D	L		
L	N	A	A	L	D	Y	M	L	T	D	-	-	-	A	S	S	A	V	R	N	R	I	D	A	S	R	L	A	V	M	G	H	S	M	G	G	G	G	T	L	R	L	A	S	Q	R	P	D	L		
L	N	A	A	L	D	Y	M	L	T	D	-	-	-	A	S	S	S	V	R	N	R	I	D	A	S	R	L	A	V	M	G	H	S	M	G	G	G	G	T	L	R	L	A	S	Q	R	P	D	L		
L	L	A	A	L	D	Y	L	T	A	K	-	-	-	-	-	S	K	V	K	D	R	I	D	P	N	R	L	A	A	M	G	H	S	M	G	G	G	G	A	L	S	A	A	V	K	R	P	S	L		
I	N	N	A	L	D	Y	L	V	S	Q	N	S	R	S	S	S	P	V	R	G	M	I	D	T	N	R	L	G	V	I	G	W	S	M	G	G	G	G	T	L	R	V	A	-	S	E	G	R	I		
I	S	A	A	L	D	Y	L	V	E	D	-	-	-	-	-	S	T	S	A	V	R	N	R	I	D	P	D	R	L	A	A	M	G	H	S	M	G	G	G	G	T	L	A	V	A	A	D	R	P	E	L
L	L	A	A	L	D	Y	L	V	Q	R	-	-	-	-	-	S	S	V	R	G	R	I	D	A	G	R	L	A	V	A	G	H	S	M	G	G	G	G	T	L	E	A	A	A	R	P	S	L			
L	L	A	A	L	D	Y	L	V	E	R	-	-	-	-	-	S	S	V	R	T	R	I	D	G	T	R	L	A	V	A	G	H	S	M	G	G	G	G	S	L	E	A	A	A	S	R	P	S	L		
L	L	A	A	L	D	Y	L	T	Q	R	-	-	-	-	-	S	S	V	R	S	R	I	D	S	S	R	L	A	V	M	G	H	S	M	G	G	G	G	S	L	E	A	S	E	S	R	P	S	L		
L	L	A	A	L	D	Y	L	T	E	R	-	-	-	-	-	S	S	V	R	G	R	V	D	A	G	R	L	A	V	A	G	H	S	M	G	G	G	G	S	L	E	A	A	A	S	R	P	S	L		
L	R	A	A	L	E	Y	L	T	E	R	-	-	-	-	-	S	D	V	A	D	R	V	D	P	D	R	L	A	V	M	G	H	S	M	G	G	G	G	T	L	E	A	A	A	K	D	D	P	E	L	
L	I	E	A	A	L	D	Y	L	V	E	D	-	-	-	-	-	S	D	V	A	D	R	V	D	G	N	R	L	A	V	M	G	H	S	M	G	G	G	G	T	L	E	A	A	A	E	N	R	P	E	L
L	L	A	A	L	D	Y	L	T	Q	R	-	-	-	-	-	S	D	V	R	S	R	V	D	A	T	R	L	G	V	M	G	H	S	M	G	G	G	G	S	L	R	A	A	A	Q	R	P	A	L		
L	L	A	A	L	D	Y	L	T	E	R	-	-	-	-	-	S	D	V	A	D	R	V	D	P	S	R	L	A	V	M	G	H	S	M	G	G	G	G	T	L	E	A	A	K	D	R	P	S	L		
L	L	A	A	V	D	Y	L	A	N	T	-	-	-	-	-	S	T	V	R	T	R	V	D	R	N	R	L	A	V	M	G	H	S	M	G	G	G	G	S	L	S	A	A	N	A	R	P	S	L		
L	E	S	A	L	D	Y	L	A	D	R	-	-	-	-	-	S	A	V	S	D	R	V	D	G	D	R	M	A	V	M	G	H	S	M	G	G	G	G	T	L	A	A	V	D	S	R	P	E	L		
L	L	A	A	L	D	W	A	V	S	S	-	-	-	-	-	S	P	A	A	S	R	I	D	R	N	R	L	A	V	A	G	H	S	M	G	G	G	G	T	L	E	A	T	L	D	R	P	S	L		
L	L	A	A	L	D	Y	L	T	G	K	-	-	-	-	-	S	A	V	R	D	R	I	D	A	T	R	L	G	V	M	G	H	S	M	G	G	G	G	S	L	G	A	A	A	K	R	P	A	I		
L	L	A	A	L	D	Y	L	T	G	K	-	-	-	-	-	A	S	S	A	V	A	S	R	L	D	E	S	R	I	A	V	S	G	H	S	M	G	G	G	G	T	L	E	A	A	K	A	N	S	S	L
L	L	R	A	L	D	W	A	S	S	S	-	-	-	-	-	A	P	A	A	V	R	D	R	V	D	A	T	R	Q	G	V	S	G	H	S	M	G	G	G	G	T	L	S	A	M	D	Q	R	P	S	V
L	L	A	A	L	D	Y	L	T	Q	R	-	-	-	-	-	S	A	V	R	D	R	V	D	A	S	R	L	A	V	A	G	H	S	M	G	G	G	G	S	L	E	A	A	K	A	R	T	S	L		
L	L	A	A	L	D	Y	L	T	Q	R	-	-	-	-	-	S	S	V	R	D	R	I	D	P	N	R	L	A	V	A	G	H	S	M	G	G	G	G	S	L	E	A	A	D	D	R	P	S	L		
L	L	A	A	L	D	Y	L	T	Q	R	-	-	-	-	-	S	S	V	R	G	R	V	D	A	S	R	L	A	V	A	G	H	S	M	G	G	G	G	S	L	E	A	A	D	S	R	T	S	L		
L	L	A	A	L	D	Y	L	T	Q	R	-	-	-	-	-	S	S	V	R	S	R	V	D	A	T	R	L	A	V	A	G	H	S	M	G	G	G	G	S	L	E	A	A	K	D	R	T	S	L		
L	L	A	A	L	D	Y	L	T	Q	R	-	-	-	-	-	S	S	V	R	S	R	V	D	A	S	R	L	A	V	A	G	H	S	M	G	G	G	G	S	L	E	A	A	K	S	R	T	S	L		
M	K	A	A	L	D	W	L	V	K	E	-	-	-	-	-	S	P	T	A	V	R	Q	R	I	D	P	T	R	L	G	V	M	G	W	S	M	G	G	G	G	T	L	D	A	A	E	T	H	P	N	I
L	L	A	A	L	D	Y	L	T	T	Q	-	-	-	-	-	S	P	T	A	V	R	Q	R	L	D	N	T	R	L	S	V	I	G	H	S	M	G	G	G	G	A	L	R	A	A	N	D	R	P	D	L
L	L	A	A	L	D	H	L	T	T	Q	-	-	-	-	-	S	R	V	R	D	R	I	D	P	T	R	L	A	V	L	G	H	S	M	G	G	G	A	S	L	A	A	A	A	E	R	P	T	L		
L	L	A	A	L	D	H	L	V	G	K	-	-	-	-	-	S	P	A	Q	V	R	Q	R	L	D	A	Q	R	L	G	V	M	G	H	S	M	G	G	G	A	M	Y	A	A	W	S	R	S	S	L	
L	L	A	A	L	D	Y	L	T	T	R	-	-	-	-	-	S	Q	V	R	D	R	I	D	P	S	R	L	A	V	M	G	H	S	M	G	G	G	G	S	L	A	A	A	A	K	R	P	T	L		
L	L	A	A	L	D	Y	L	T	K	S	-	-	-	-	-	S	S	V	R	G	R	I	D	S	T	R	L	G	V	M	G	H	S	M	G	G	G	G	S	L	E	A	A	K	T	R	P	S	L		
L	L	A	A	L	D	Y	L	A	E	D	-	-	-	-	-	S	D	R	D	V	R	D	R	L	D	P	D	R	L	A	V	M	G	H	S	M	G	G	G	G	S	L	E	A	T	S	D	R	P	S	L
L	L	A	A	L	D	Y	L	T	Q	Q	-	-	-	-	-	S	S	V	R	S	R	I	D	A	T	R	L	G	V	M	G	H	S	M	G	G	G	G	T	L	R	A	A	S	Q	R	P	T	L		
L	L	A	A	L	D	Y	L	T	E	R	-	-	-	-	-	S	S	V	R	S	R	I	D	S	S	R	L	G	V	M	G	H	S	M	G	G	G	G	T	L	E	A	A	K	T	R	P	S	L		
L	L	A	A	L	D	H	L	T	E	R	-	-	-	-	-	S	P	V	R	S	R	V	D	S	D	R	L	A	V	M	G	H	S	M	G	G	G	G	S	L	E	A	A	K	S	R	P	S	L		
L	L	A	A	L	D	Y	L	T	Q	R	-	-	-	-	-	S	S	V	R	G	R	I	D	A	S	R	L	A	V	A	G	H	S	M	G	G	G	G	S	L	E	A	A	A	S	R	P	S	L		
L	L	A	A	L	D	Y	L	T	R	S	-	-	-	-	-	S	S	V	R	S	R	I	D	A	S	R	L	A	V	S	G	H	S	M	G	G	G	G	S	L	E	A	A	S	D	R	P	S	L		
L	L	A	A	L	D	Y	L	T	R	S	-	-	-	-	-	S	S	V	R	S	R	I	D	S																											

Catalytic Aspartate

K	A	A	I	P	L	T	P	W	N	L	D	K	T	W	S	E	V	R	V	P	T	L	I	I	G	A	E	N	D	T	I	A	P	V	A	X	H	S	E	P	F	F	Y	N	S	L	P	S	S	S	L
K	A	A	I	P	L	T	P	W	H	L	N	K	N	W	S	S	V	T	V	P	T	L	I	I	G	A	D	L	D	T	I	A	P	V	A	T	H	A	K	P	F	F	Y	N	S	L	P	S	S	S	I
K	A	A	A	A	P	Q	A	P	W	D	S	S	T	N	F	S	S	V	T	V	P	T	L	I	F	A	C	E	N	D	S	I	A	P	V	N	S	S	A	L	P	I	Y	D	S	M	-	S	R	N	
K	A	A	V	P	L	T	P	W	H	T	D	K	T	F	N	-	T	S	V	P	V	L	I	V	G	A	E	A	D	T	V	A	P	V	S	Q	H	A	I	P	F	Y	Q	N	L	P	S	T	T		
K	A	S	I	P	L	T	P	W	N	L	D	K	T	W	G	Q	V	Q	V	P	T	F	I	I	G	A	E	L	D	T	I	A	S	V	R	T	H	A	K	P	F	F	Y	E	S	L	P	S	S	L	
K	A	A	I	P	L	T	P	W	H	L	N	K	N	W	S	S	V	R	V	P	T	L	I	I	G	A	D	L	D	T	I	A	P	V	L	T	H	A	R	P	F	F	Y	N	S	L	P	T	S	I	
K	A	A	I	P	L	T	P	W	H	L	N	K	S	W	R	D	I	T	V	P	T	L	I	I	G	A	E	Y	D	T	I	A	S	V	T	L	H	S	K	P	F	F	Y	N	S	I	P	S	P	T	
K	A	A	I	P	L	T	G	W	N	T	D	K	T	W	P	E	L	R	T	P	T	L	V	V	G	A	D	G	D	T	V	A	P	V	A	T	H	S	K	P	F	F	Y	E	S	L	P	G	S	L	
K	A	A	I	P	L	T	P	W	H	L	N	K	N	W	S	S	V	R	T	P	T	L	I	I	G	A	E	L	D	S	I	A	P	V	A	T	H	S	I	P	F	F	Y	N	S	L	T	N	A	R	
K	A	A	I	P	L	T	P	W	H	L	N	K	N	W	S	S	V	R	V	P	T	L	I	I	G	A	D	L	D	T	I	A	P	V	L	T	H	A	R	P	F	F	Y	N	S	L	P	T	S	I	
Q	A	A	I	P	L	T	P	W	H	T	D	K	T	W	G	S	V	R	V	P	T	L	I	I	G	A	E	N	D	T	I	A	S	V	R	S	H	S	E	P	F	F	Y	N	S	L	P	G	S	L	
K	A	A	I	P	L	T	P	W	H	L	N	K	S	W	R	D	I	T	V	P	T	L	I	I	G	A	E	Y	D	T	I	A	S	V	T	L	H	S	K	P	F	F	Y	N	S	I	P	S	P	T	
K	A	A	I	P	L	T	P	W	H	L	N	K	S	W	R	D	I	T	V	P	T	L	I	I	G	A	D	L	D	T	I	A	P	V	S	S	H	S	E	P	F	F	Y	N	S	I	P	S	S	T	
K	A	V	V	P	L	A	P	W	Y	V	G	G	L	E	Q	S	T	V	P	T	M	I	F	G	A	D	N	D	F	I	A	P	V	A	S	N	A	R	P	F	F	Y	Q	S	L	T	K	V	P		
K	A	A	I	P	L	A	P	W	D	T	T	S	Y	Y	S	R	S	Q	A	P	T	L	I	F	A	C	E	S	D	V	I	A	P	V	L	Q	H	A	S	P	F	F	Y	N	S	L	P	S	S	I	
K	A	A	I	P	L	T	P	W	N	L	D	K	T	W	S	S	V	D	V	P	T	M	I	I	G	A	E	N	D	T	V	A	S	V	S	S	H	S	I	P	F	F	Y	N	S	L	S	G	A	P	
Q	A	A	V	P	L	A	P	W	N	L	D	K	T	W	S	G	V	T	V	P	T	L	I	V	G	G	E	S	D	S	V	A	P	V	A	S	H	S	V	P	F	F	Y	N	S	I	P	A	S	S	
Q	A	A	V	P	L	A	P	W	N	L	D	K	T	W	N	T	L	Q	V	P	T	L	I	I	G	G	E	N	D	S	V	A	P	V	A	T	H	S	I	P	F	F	Y	N	S	I	P	A	S	A	
K	A	S	I	P	L	T	P	W	N	L	K	K	N	W	R	N	N	R	V	P	T	L	I	V	G	A	E	N	D	S	I	A	S	V	R	T	H	A	E	P	F	F	Y	E	S	I	P	S	T	T	
Q	A	A	V	P	L	A	P	W	N	L	S	K	N	W	S	S	L	R	V	P	T	L	I	I	G	G	E	S	D	S	V	A	P	V	S	S	H	S	V	P	F	F	Y	N	S	I	P	A	S	A	
K	A	A	I	P	L	T	P	W	N	T	D	K	T	W	P	E	V	S	T	P	T	L	V	V	G	A	Q	D	D	S	I	A	S	V	T	H	A	E	P	F	F	Y	E	S	L	P	G	A	T		
R	A	A	I	P	L	T	P	W	H	L	Q	K	N	W	S	D	V	E	V	P	T	M	I	I	G	A	E	N	D	T	V	A	S	V	R	T	H	S	I	P	F	F	Y	E	S	L	D	E	D	L	
Q	A	A	I	P	L	T	P	W	H	T	T	K	N	W	S	T	V	R	V	P	T	M	V	I	G	A	E	N	D	S	I	A	P	V	A	Q	H	S	E	P	F	F	Y	T	S	L	P	S	T	L	
K	A	I	I	P	M	T	P	W	N	L	D	K	T	W	G	D	V	E	V	P	M	L	D	F	G	A	E	Y	D	T	I	A	S	P	T	T	H	A	K	P	F	F	Y	N	T	-	-	H	G	G	
Q	A	A	I	P	L	T	G	W	H	T	V	K	N	W	S	S	V	R	V	P	T	L	V	V	G	A	E	N	D	T	V	A	P	V	S	S	H	S	E	P	F	F	Y	T	S	L	P	A	T	L	
R	A	A	I	P	L	T	P	W	H	T	Q	K	N	W	S	S	V	Q	T	P	T	L	V	V	G	A	Q	N	D	V	T	A	S	V	R	S	H	A	I	P	M	Y	E	S	L	P	S	S	L		
K	A	A	I	P	L	A	P	W	N	L	T	K	S	W	T	R	Q	R	V	P	T	V	I	I	G	G	S	S	D	T	I	A	P	V	S	S	H	S	E	P	F	F	Y	T	S	L	-	G	S	P	
Q	A	A	V	P	L	T	P	W	H	L	D	K	T	W	N	E	V	R	V	P	T	M	I	I	G	A	E	A	D	T	T	A	S	V	G	S	H	S	E	R	F	F	Y	A	S	L	T	N	A	R	
K	A	A	V	A	L	Q	P	W	H	T	D	K	T	W	P	E	V	R	V	P	T	M	I	I	G	A	E	D	D	T	I	A	P	V	A	S	H	S	I	P	F	F	Y	T	S	A	T	G	A	R	
R	A	G	V	P	L	A	P	W	H	T	T	S	W	P	R	V	T	N	P	V	M	I	L	G	Q	N	D	G	I	A	P	V	S	S	H	A	I	P	M	Y	T	G	V	A	S	G	-				
K	A	A	I	P	L	A	P	W	N	L	D	K	T	W	P	E	V	R	T	P	T	L	I	I	G	G	E	L	D	A	V	A	P	V	A	T	H	S	I	P	F	F	Y	N	S	L	S	N	A	P	
Q	A	A	V	P	I	A	P	W	N	L	T	K	S	W	S	G	L	R	V	P	T	M	I	V	G	G	E	S	D	S	I	A	P	V	S	S	H	S	I	P	F	F	Y	T	S	I	P	A	S	A	
Q	A	A	V	P	L	A	P	W	H	T	T	K	N	W	S	G	V	R	V	P	T	M	I	I	G	G	E	S	D	S	V	A	P	V	S	T	H	S	E	L	F	Y	A	S	I	P	A	S	S		
K	A	A	I	P	L	A	P	W	N	T	D	K	T	W	P	E	V	R	T	P	T	L	I	I	G	G	E	S	D	T	V	A	P	V	S	S	H	S	I	P	F	F	Y	T	S	L	T	S	A	R	
K	A	A	I	P	L	A	P	W	N	L	D	K	T	W	P	E	V	T	P	T	M	I	F	G	G	E	L	D	T	V	A	P	V	S	T	H	A	I	P	F	F	Y	N	S	L	T	N	A	R		
Q	A	A	I	P	L	A	P	W	H	T	D	K	D	W	S	T	V	K	V	P	T	M	I	I	G	A	Q	N	D	F	I	A	G	V	G	A	H	S	E	P	F	F	Y	E	S	L	K	S	A	P	
Q	A	A	V	P	L	G	P	Y	H	T	T	K	D	W	S	G	V	R	V	P	T	M	I	I	G	G	Q	Y	D	S	I	A	P	N	D	Q	H	S	E	R	F	F	Y	N	S	L	D	S	A	P	
Q	A	A	V	P	L	A	P	W	H	T	D	K	T	W	N	D	V	R	V	P	T	M	I	I	G	G	E	N	D	N	I	A	P	V	R	T	H	A	E	A	F	F	Y	T	S	M	S	N	A	P	
K	A	A	V	P	L	A	P	Y	H	T	I	K	N	W	S	A	I	Y	V	P	T	L	F	F	A	G	T	E	D	T	V	T	R	P	E	D	H	A	E	R	F	F	Y	S	M	I	Y	A	R		
R	A	A	I	P	L	A	P	W	S	L	T	K	N	W	S	D	L	T	V	P	T	L	I	I	G	A	E	N	D	N	V	A	P	V	A	G	H	S	E	R	F	F	Y	D	S	M	T	N	V	P	
Q	A	A	I	P	L	T	P	W	N	L	D	K	T	W	P	E	I	K	T	P	T	L	I	F	G	A	D	G	D	T	I	A	P	V	A	T	H	A	E	P	F	F	Y	N	T	L	P	S	S	L	
K	A	S	I	P	L	T	P	W	N	L	D	K	T	W	Y	E	V	T	V	P	T	F	I	I	G	A	E	S	D	S	I	A	S	V	R	T	H	S	E	P	F	F	Y	E	S	L	S	G	S	L	
Q	A	A	I	P	L	T	A	W	H	T	T	K	N	W	S	S	V	R	V	P	T	L	V	V	G	A	E	D	D	S	I	A	P	V	A	T	H	S	E	P	F	F	Y	T	L	P	S	T	L		
Q	A	A	I	P	L	T	P	W	N	L	D	K	T	W	P	E	V	R	T	P	T	L	I	I	G	A	D	G	D	T	I	A	P	V	A	S	H	S	E	P	F	F	Y	S	S	L	P	S	S	T	
K	A	A	I	P	L	T	P	W	N	L	D	K	T	W	P	E	L	R	T	P	T	L	I	V	G	A	E	R	D	S	I	A	P	V	S	S	H	A	E	P	F	F	Y	N	T	L	P	S	S	L	
Q	A	A	V	P	L	A	P	W	N	L	D	K	T	W	S	E	L	R	V	P	T	L	I	V	G	G	E	A	D	S	V	A	P	V	A	T	H	S	E	P	F	F	Y	N	S	I	P	A	S	S	
Q	A	A	V	P	L	A	P	W	N	T	D	K	L	W	T	E	L	R	V	P	T	L	I	V	G	Q	A	D	A	I	A	P	V	A	T	H	S	I	P	F											

T F G A V A I S P G Y T A R Q S S I A W L G P R L A S H G F V V I T I D T N T T L D Q P D S R G R Q
 T V G A I A I V P G Y T A R Q S S I K W W G P R L A S H G F V V I T I D T N S T L D Q P D S R S S Q
 T Y G A V A I S P G Y T G T E A S I A W L G E R I A S H G F V V I T I D T I T T L D Q P D S R A E Q
 T F G A V A I S P G F T A S Q G S M S W Y G E R V A S H G F V V I F T I D T N T R L D Q P G O R G R Q
 T Y G A V A I S P G Y T G T Q A S V A W L G E R I A S H G F V V I T I D T N T T L D Q P D S R A R Q
 T F G G I A M S P G Y T A D A S S L A W L G E R R L A S H G F V V L V I N T N S R F D Y P D S R A S Q
 T Y G A I A I S P G Y T G T Q S S I A W L G E R I A S H G F V V I A I D T N T T L D Q P D S R A R Q
 T Y G A I A I S P G Y T G T Q S S I A W L G E R I A S H G F V V I A I D T N T T L D Q P D S R A R Q
 T Y G A I A I S P G Y T G T Q S S I A W L G E R I A S H G F V V I A I D T N T T L D Q P D S R A R Q
 T Y G A I A I S P G Y T G T Q S S I A W L G E R I A S H G F V V I A I D T N T T L D Q P D S R A R Q
 K M G A I A V V I P G F T A G E S S I A W W G P R L A S H G F V V I T I A T N S G F D Q P A S R A T Q
 T M G A I A V I P G Y V S Y E S S I K W W G P R L A S W G F V V I T T D T N T I Y D Q P D S R A D Q
 O Y G V V A L S P G F T A T Q S S V A W L G R R I A T H G F V V V T I N T N S T L D Q P A S R A N Q
 T M G A I A V V I P G F F A S A E S S I A W W G P K L A S H G F V V M T I D T N S G F D Q P P S R A T Q
 N M G A I A V V I P G F V S Y E S S I E W W G P R L A S W G F V V I T I D T N T T V D Q P D S R A D Q
 T M G A I A V V I P G F L L Q E S S I D F W G P K L A S H G F V V I T I S A N S G F D Q P A S R A T Q
 T M G A I A V V I P G F V S A E S S I E W W G P K L A S H G F V V M T I D T N S G F D Q P P S R A N Q
 K M G A I A V I P G Y V S Y E S S I R W W G S R L A S W G F V V I T I D T N T I Y D Q P D S R A D Q
 N M G A I A V I P G Y V S Y E S S I E W W G P R L A S W G F V V I T I D T N T I Y D Q P D S R A D Q
 O Y G A I A V V I P G Y T A S K S T V T F W G E R L S S H G F V V I V I D T L T T L D Q P A S R A R Q
 S V G A V A V V I P G Y L A R Q S S I R W W G P R L A S W G F V V I T L D T R S T S D Q P A S R A R Q
 T V G G I V V V P G Y T A Y Q S D I N W W G P R L A S W G F V V M T I D T N S T L D Q P A S R S R Q
 K V G A I A V V I P G Y T A R Q S S I N W W G P R L A S H G F V V I T I D T N S T L D Q P A S R S S Q
 T V G A I A I V P G F L A R Q S S I N W W G P R L A S H G F V V I T I D T N S T F D Q P T S R S R Q
 T V G A I A I V P G F L A R Q S S I N W W G P R L A S H G F V V I T I D T N S T F D Q P T S R S R Q
 T F G A V A I S P G F V S P K S W I D W Y G P R L A S O G F V V M T L E T F S Y F D A P D G R A D Q
 T M G A V V I P G F V S A E S S I D W W G P K L A S Y G F V V M T I D T N T G F D Q P P S R A R Q
 T F G A V I A P G Y T A S S S S M S W M G E R L A S O G F V I F T I D T N T R Y D Q P P S R A D Q
 T F G A I A I S P G Y T A T W S S L S W L G P R I A S H G F V V I G I E T N T L L D Q P D S R G R Q
 T F G A I A I S P G Y T A W S S I S W L G P R I A S H G F V V I G I E T N S R L D Q P D S R G R Q
 T F G A V A I A P G F T A S E G S M A W Y G P R I A S O G F V V F T I D T L T R L D Q P D S R G R Q
 T F G A V A I S P G Y T A S S S M D W L G P R I A S H G F V V I G I D T N S R L D Q P A S R G R Q
 T F G A V A V A P G Y T A T E S S M A W L G P R L A S O G F V V F T I N T R T T S D Q P D S R G D Q
 T F G G V I A P G Y T A S Q S S M A W M G H R I A S O G F V V F T I D T I T R Y D Q P P S R G D Q
 T F G A V A I S P G F T G T Q A T I A W L G P R I A S O G F V V F T I D T T T V F D Q P D S R A T Q
 T F G A V A V V I P G Y T A S S T I E W L G A R V A S O G F V V F T I D T D T R L D Q P R Q R G T Q
 T F G A V A I S P G F T A S S Q S S V A W L G P R L A S O G F V V I T I D T L S V Y D Q P A S R G T Q
 T F G G V V V A P G Y T A S E S S M S W Y G H R F A S O G F V V F T I D T N T R Y D Q P N S R G R Q
 T F G A V A I S P G F T A Y W S S L D W L G P R L A S O G F V V I G I E T N T T L D Q P A Q R G T Q
 T F G A I A V I P G F T G P E S S I A W F G P H L A S H G F V V M T L T T S S G W D F P A S R D N Q
 T Y G G V V L A P G F T A S S W M Y D S I A R R A A S H G F V V F A I D T N S I Y D Q P G S R G T Q
 R F P V A I S P G Y T E R W S S F A W L G R R L A S W G F V V V G I E T N S L F D Q P N S R G T Q
 T F G A I A I S P G F T A Y W S S L E W L G H R L A S O G F V V I G I E T N T T L D Q P D O R G T Q
 T F G A V A I S P G Y T A R W S S L D W L G P R I A S O G F V V I G I E T N T L Y D Q P S S R G R Q
 T F G G I A I S P G Y T A A W S S L D W L G P R I A S H G F V V I G I D T L S R Y D Q P A S R G R Q
 T F G G I A I S P G Y T A A W S S L D W L G P R L A S H G F V V I G I E T N T R L D Q P A S R G D Q
 T F G A V A I S P G F T A D W T S L D W L G P R L A S H G F V V I G I D T I T R L D Q P D S R G R Q
 T F G A I A I T P G F L S P E S S I A W L G P R L A S N G F V V I T L T T N T G F D Q P N G R A D Q
 T F G G V A V S P G F L S P E I L I A W L G P R L A S H G F V V I T F T N T P T D Q P D S R G S Q
 T F G A I A A I P G F V S P E A I V R W F G P R L A S O G F V V I T L T S D G A T D T P A S R G D Q
 T F G G V A I S P G Y T G P E G H V A W L G P R L A S O G F V V F T I A T N S V Y D Q P T E R G O Q
 T F G A V A V I P G F L S P E S L V R W F G P R L A S O G F V V M T L T T N G L T D T P E S R S E Q
 T F G A V A V S P G Y T G T Q S S I A W L G P R L A S O G F V V F T I D T L T T L D Q P D S R G N Q
 T F G G V A I S P G Y T A G E S S M S W Y G P R L A S O G F V V F T I D T N T R F D Q P G S R A S Q
 T F G A V A V S P G Y T G T Q S S I S W L G P R L A S O G F V V F T I D T N T I Y D Q P D S R A S Q
 T F G A V I S P G F T A L Q S S I A W L G P R L A S O G F V V F T I D T N T T V D Q P D S R G R Q
 T F G A I A I S P G Y T A R E S S I A W L G P R L A S O G F V V F T I D T N S V Y D Q P A S R G R Q
 T F G A I A I S P G Y T A A W S S I D W L G P R L A S H G F V V I G I E T I T R L D Q P D S R G R Q
 T F G A I A I S P G F T A Y W S S I S W L G P R L A S H G F V V I G I E T L T T L D Q P D S R G Q Q
 T F G A V A I S P G F T A T W S S L S W L G P R I A S H G F V V I G I E T N T L L D Q P G O R G D Q
 T F G A I A I S P G Y T A S W S S I S W L G P R I A S H G F V V I G I E T N T R L D Q P D S R G R Q
 T F G A V A I S P G Y T A T Q S S I A W L G P R L A S O G F V V L T I D T L T T L D Q P D S R G R Q
 T F G A V A V A P G Y T A S Q S S L S W L G P R L A S O G F V V F T I D T L S T L D Q P D A R G R Q
 T F G A V I S P G F T A Y Q S S I A W L G P R L A S O G F V V F T I D T N T T L D Q P D S R G R Q
 T F G A I A I S P G F T A R W S S I D W L G P R I A S H G F V V I G I E T N S V Y D Q P S S R G T Q
 T F G A V A I S P G F T A T W S S L A W L G P R I A S H G F V V I G I E T N T L L D Q P D S R G D Q
 T F G G V A I A P G F T A G E G T M S W Y G P R L A S O G F V V F T I D T N T R L D Q P S A G R G A Q
 T F G A V A I A P G Y T S L Q S S I A W L G P R L A S O G F V V I T I D T R T T L D Q P A G R G D Q
 T F G A L A I S P G F T A T Q S S I S W M G P R L A S O G F V V I T I D T L S T L D Q P D S R G D Q
 T F G A V A I S P G F T G T Q S S L S W L G P R L A S O G F V V F T I D T I T V L D Q P A S R G H Q
 T F G A V A V S P G Y T A S Q T S M S W Y G P R L A S O G F V V F T I D T N S R Y D Q P G S R G D Q
 T Y G A L A I S P G F T A T Q S S I S W M G R R L A S O G F V V I T I D T N T I Y D Q P A S R G D Q
 T F G A V A V S P G F T A S Q S S V A W Y G P R L A S O G F V V M T I D T L S V Y D Q P A S R G D Q
 T F G A V A I S P G F T A Y Q S S I A W L G P R L A S R G F V V M T I D T N S T L D Q P A S R G D Q

50 60 70 80 90 100

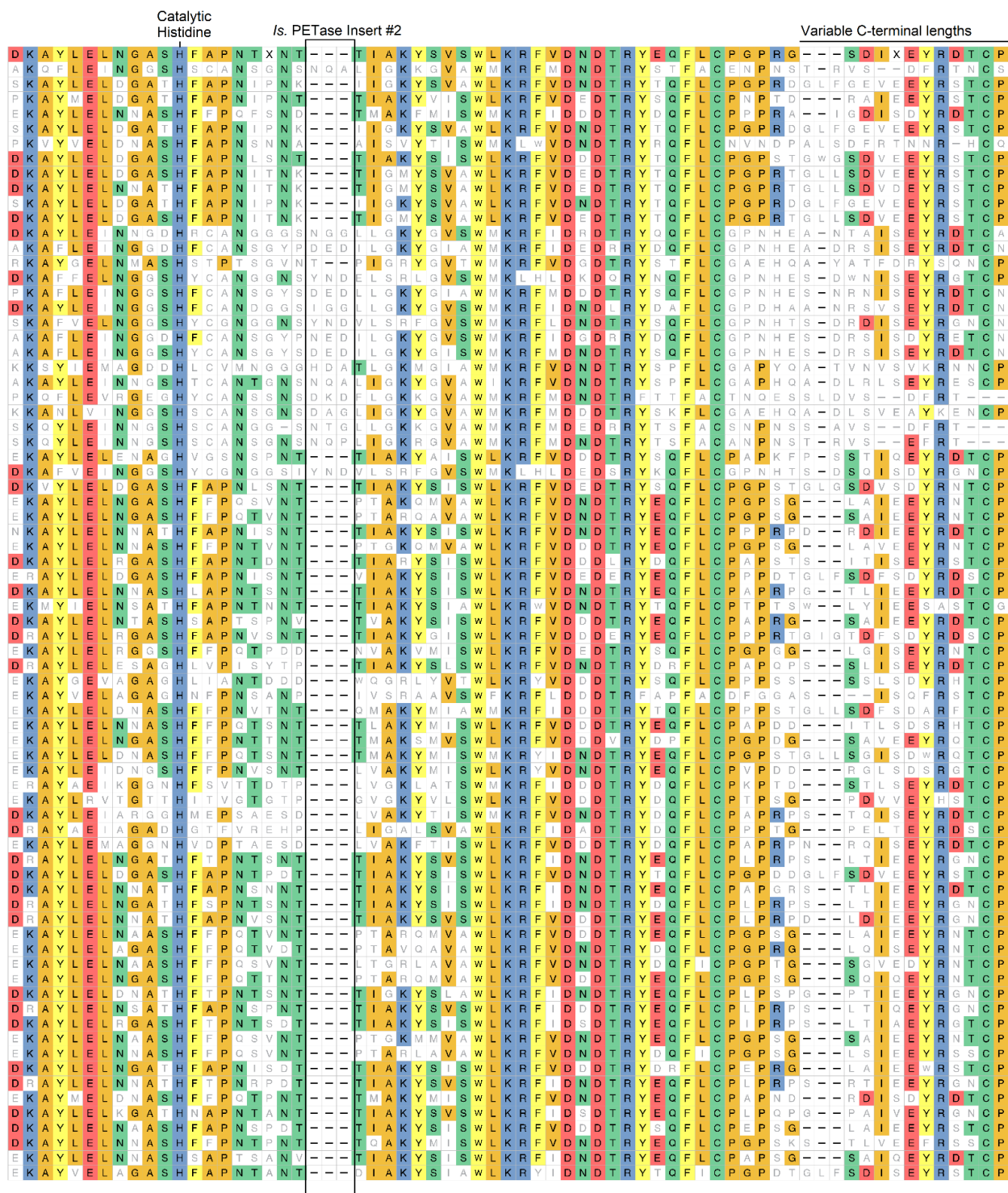
/s. PETase Insert #1

Catalytic Serine

L	L	A	A	L	D	Y	L	T	Q	---	R	---	S	S	V	R	S	R	I	D	A	S	R	L	A	V	M	G	H	S	M	G	G	G	G	S	L	E	A	A	X	D	R	P	S	L				
Q	M	A	A	L	R	Q	V	A	S	L	N	G	T	S	S	S	P	I	Y	G	K	V	D	T	A	R	M	G	V	M	G	H	S	M	G	G	G	G	S	L	I	S	A	A	N	N	P	S	L	
L	L	A	A	L	N	H	M	I	N	---	R	A	S	S	T	V	R	S	R	I	D	S	S	R	L	A	V	M	G	H	S	M	G	G	G	G	S	L	R	L	A	S	Q	R	P	S	L			
L	L	A	A	L	D	Y	L	T	Q	---	R	---	S	S	V	R	N	R	V	D	A	S	R	L	A	V	A	G	H	S	M	G	G	G	G	T	L	E	A	A	K	S	R	T	S	L				
L	N	A	A	L	D	Y	M	I	N	---	D	A	S	S	A	V	R	S	R	I	D	S	S	R	L	A	V	M	G	H	S	M	G	G	G	G	T	L	R	L	A	S	Q	R	P	S	L			
L	N	A	A	L	N	Y	L	R	T	---	S	S	P	S	A	V	R	A	R	L	D	A	N	R	L	A	V	A	G	H	S	M	G	G	G	G	T	L	R	I	A	E	Q	N	P	S	L			
L	D	A	A	L	D	Y	M	V	E	---	D	S	S	Y	S	V	R	N	R	I	D	S	S	R	L	A	A	M	G	H	S	M	G	G	G	T	L	R	L	A	E	R	P	D	L					
L	N	A	A	L	D	Y	M	L	T	---	D	A	S	S	A	V	R	N	R	I	D	A	S	R	L	A	V	M	G	H	S	M	G	G	G	T	L	R	L	A	S	Q	R	P	D	L				
L	N	A	A	L	D	Y	M	I	N	---	D	A	S	S	A	V	R	S	R	I	D	S	S	R	L	A	V	M	G	H	S	M	G	G	G	S	L	R	L	A	S	Q	R	P	D	L				
L	N	A	A	L	D	Y	M	L	T	---	D	A	S	S	A	V	R	N	R	I	D	A	S	R	L	A	V	M	G	H	S	M	G	G	G	T	L	R	L	A	S	Q	R	P	D	L				
L	S	N	A	L	D	Y	V	I	S	---	Q	S	N	S	S	S	A	I	S	G	M	V	D	A	N	R	L	G	A	M	G	H	S	M	G	G	G	T	L	R	L	A	S	G	A	-	R	L		
L	S	A	A	L	D	Y	V	I	S	---	Q	S	N	S	S	S	P	I	Y	G	M	V	D	A	N	R	L	G	A	M	G	H	S	M	G	G	G	T	L	K	L	S	T	E	R	-	E	L		
L	I	A	A	L	N	Y	V	A	N	---	S	A	S	T	T	V	R	S	R	V	D	A	S	R	L	A	V	G	H	S	M	G	G	G	S	L	I	A	A	Q	N	N	P	S	L					
T	N	N	A	L	D	Y	L	I	D	---	Q	S	N	S	S	S	P	I	N	G	M	V	D	T	S	R	L	G	A	V	G	H	S	M	G	G	G	T	L	R	V	A	A	E	G	-	R	L		
L	S	A	A	L	D	L	L	V	S	---	S	G	S	S	A	I	S	G	L	V	D	P	S	R	L	G	A	I	G	H	S	M	G	G	G	S	L	K	L	A	T	E	R	-	N	L				
L	G	R	A	L	D	Y	V	I	N	---	Q	S	N	G	S	N	S	P	I	S	G	M	V	D	T	T	R	L	G	V	V	G	H	S	M	G	G	G	A	L	Q	L	A	S	G	D	-	R	L	
T	N	A	A	L	D	Y	L	V	A	---	Q	N	R	S	S	R	S	P	V	R	G	M	I	D	T	N	R	L	G	V	V	G	H	S	M	G	G	G	T	L	R	V	A	S	E	G	-	R	I	
L	S	A	A	L	D	H	L	I	D	---	Q	S	N	S	R	N	S	I	S	G	M	V	D	S	N	R	L	G	V	I	G	H	S	M	G	G	G	T	L	K	L	A	T	E	R	-	N	L		
L	K	A	A	L	D	K	V	V	A	---	L	S	K	T	S	S	S	P	I	Y	N	K	V	D	P	S	R	L	A	V	M	G	H	S	M	G	G	G	T	L	Q	A	A	K	D	N	P	S	L	
Q	M	A	A	L	D	Q	L	A	T	---	L	S	R	T	S	S	P	I	Y	G	K	V	D	P	N	R	L	A	V	M	G	H	S	M	G	G	G	T	L	I	S	A	R	D	N	P	S	L		
Q	M	A	A	L	R	Q	V	A	T	---	L	A	G	T	S	S	S	P	I	Y	N	K	V	D	T	A	R	L	G	V	M	G	H	S	M	G	G	G	S	L	I	S	A	K	N	N	P	S	L	
Q	M	A	A	L	R	Q	V	A	T	---	L	A	G	T	S	S	S	P	I	Y	R	K	V	D	T	A	R	L	G	V	M	G	H	S	M	G	G	G	T	L	L	A	A	R	D	N	P	S	L	
L	L	A	A	L	D	Q	V	A	A	---	L	A	R	T	S	S	S	P	I	Y	N	R	K	V	D	T	T	R	M	G	V	M	G	H	S	M	G	G	G	T	L	I	S	A	Q	N	R	P	S	L
L	L	A	A	L	D	Y	L	T	A	---	K	---	S	K	V	K	D	R	I	D	P	N	R	L	A	A	M	G	H	S	M	G	G	G	A	L	S	A	A	V	K	R	P	S	L					
T	N	N	A	L	D	Y	L	V	S	---	Q	N	S	R	S	S	P	V	R	G	M	I	D	T	N	R	L	G	V	I	G	H	S	M	G	G	G	T	L	R	V	A	A	S	E	G	-	R	I	
T	S	A	A	L	D	Y	L	V	E	---	D	S	T	S	A	V	R	N	R	I	D	P	N	R	L	A	A	M	G	H	S	M	G	G	G	T	L	A	V	A	A	S	D	R	P	E	L			
L	L	A	A	L	D	Y	L	V	Q	---	R	---	S	S	V	R	G	R	I	D	A	G	R	L	A	V	A	G	H	S	M	G	G	G	T	L	E	A	A	A	A	D	R	P	S	L				
L	L	A	A	L	D	Y	L	V	E	---	R	---	S	S	V	R	T	S	R	I	D	G	T	R	L	A	V	A	G	H	S	M	G	G	G	S	L	E	A	A	A	S	R	P	S	L				
L	L	A	A	L	D	Y	L	T	Q	---	R	---	S	S	V	R	S	R	I	D	S	S	R	L	A	V	M	G	H	S	M	G	G	G	S	L	E	A	A	S	E	S	R	P	S	L				
L	L	A	A	L	D	Y	L	T	E	---	R	---	S	S	V	R	G	R	V	D	A	G	R	L	A	V	A	G	H	S	M	G	G	G	S	L	E	A	A	A	S	R	P	S	L					
L	R	A	A	L	E	Y	L	T	T	---	R	---	S	D	V	A	D	R	I	D	P	D	R	L	A	V	M	G	H	S	M	G	G	G	T	L	E	A	A	K	D	D	P	S	L					
T	E	A	A	L	D	Y	L	V	E	---	D	---	S	D	V	A	D	R	V	D	G	N	T	R	L	A	V	M	G	H	S	M	G	G	G	T	L	A	A	E	N	R	P	E	L					
L	L	A	A	L	D	Y	L	T	Q	---	R	---	S	D	V	R	S	R	V	D	A	T	R	L	G	V	M	G	H	S	M	G	G	G	S	L	R	A	A	A	Q	R	P	A	L					
L	L	A	A	L	D	Y	L	T	E	---	R	---	S	D	V	A	D	R	V	D	P	S	R	L	A	V	M	G	H	S	M	G	G	G	T	L	E	A	A	K	D	R	P	S	L					
L	L	A	A	L	V	D	Y	L	A	N	---	T	---	S	T	V	R	T	R	V	D	R	N	R	L	A	V	M	G	H	S	M	G	G	G	S	L	S	A	A	N	A	S	R	P	E	L			
L	E	S	A	L	D	Y	L	A	D	---	R	---	S	A	V	S	D	R	V	D	G	N	D	R	M	A	V	M	G	H	S	M	G	G	G	T	L	A	A	V	D	S	R	P	E	L				
L	L	A	A	L	D	w	A	V	S	---	S	---	S	P	A	A	S	R	I	D	R	N	R	L	A	V	A	G	H	S	M	G	G	G	T	L	E	A	T	L	D	R	P	S	L					
L	L	A	A	L	D	Y	L	T	G	---	K	---	S	A	V	D	R	I	D	A	T	R	L	G	V	M	G	H	S	M	G	G	G	S	L	G	A	A	K	R	P	A	I							
L	L	A	A	V	D	Y	L	K	K	---	G	A	S	A	V	A	S	R	L	D	E	S	R	I	A	V	S	G	H	S	M	G	G	G	T	L	E	A	A	K	A	N	S	L						
L	L	R	A	L	D	w	A	S	S	---	S	A	P	A	A	V	D	R	V	D	A	T	R	Q	G	V	S	G	H	S	M	G	G	G	T	L	S	A	M	D	Q	R	P	S	V					
L	L	A	A	L	D	Y	L	T	Q	---	R	---	S	A	V	R	D	R	V	D	A	S	R	L	A	V	A	G	H	S	M	G	G	G	S	L	E	A	A	K	A	R	T	S	L					
L	L	A	A	L	D	Y	L	T	Q	---	R	---	S	S	V	R	D	R	I	D	P	N	R	L	A	V	A	G	H	S	M	G	G	G	S	L	E	A	A	D	D	R	P	S	L					
L	L	A	A	L	D	Y	L	T	Q	---	R	---	S	S	V	R	G	R	V	D	A	S	R	L	A	V	A	G	H	S	M	G	G	G	S	L	E	A	A	D	S	R	P	S	L					
L	L	A	A	L	D	Y	L	T	Q	---	R	---	S	S	V	R	S	R	V	D	A	T	R	L	A	V	A	G	H	S	M	G	G	G	S	L	E	A	A	K	D	R	T	S	L					
L	L	A	A	L	D	Y	L	T	Q	---	R	---	S	S	V	R	S	R	V	D	A	P	T	R	L	A	V	A	G	H	S	M	G	G	G	S	L	E	A	A	K	S	R	T	S	L				
M	K	A	A	L	D	w	L	V	K	---	E	S	P	T	A	V	R	Q	R	I	D	A	T	R	L	G	V	M	G	H	S	M	G	G	G	T	L	D	A	A	E	T	H	P	N	I				
L	L	A	A	L	D	Y	L	T	T	---	Q	S	P	T	A	V	R	Q	R	L	D	N	T	R	L	S	V	I	G	H	S	M	G	G	G	A	L	R	A	A	N	D	R	P	D	L				
L	L	A	A	L	D	H	L	T	T	---	Q	---	S	R	V	R	D	R	I	D	A	P	T	R	L	A	V	L	G	H	S	M	G	G	G	A	S	L	A	A	A	E	R	P	T	L				
L	L	A	A	L	D	H	L	V	G	---	K	S	P	A	Q	V	R	Q	R	L	D	A	Q	R	L	G	V	M	G	H	S	M	G	G	G	A	M	Y	A	A	A	E	R	P	T	L				
L	L	A	A	L	D	Y	L	T	T	---	R	---	S	Q	V	R	D	R	I	D	P	S	R	L	A	V	M	G	H	S	M	G	G	G	S	L	A	A	A	K	R	P	T	L						
L	L	A	A	L	D	Y	L	T	K	---	S	---	S	S	V	R	G	R	I	D	S	P	T	R	L	G</																								

Catalytic Aspartate

K	A	A	I	P	L	A	P	W	N	L	D	K	T	W	S	E	V	R	V	P	T	L	I	I	G	A	E	N	D	T	I	A	P	V	A	S	H	S	E	P	F	Y	N	S	L	P	S	S	L	
K	A	A	A	P	L	A	P	W	D	S	S	T	N	F	S	S	V	T	V	P	T	L	I	I	F	A	C	E	N	D	S	I	A	P	V	N	S	A	L	P	I	Y	D	S	M	S	R	N	-	
K	A	A	S	I	P	L	T	P	W	H	L	N	K	N	W	S	S	V	T	V	P	T	L	I	I	G	A	D	L	D	T	I	A	P	V	A	T	H	A	K	P	F	Y	N	S	L	P	S	S	L
K	A	A	S	I	P	L	T	P	W	N	L	D	K	T	W	P	E	V	R	T	P	T	F	I	I	G	A	E	L	D	S	I	A	S	V	R	T	H	A	K	P	F	Y	N	S	L	P	S	S	L
K	A	A	I	P	L	A	P	W	H	L	D	K	T	W	P	E	V	R	T	P	T	L	I	I	G	A	E	L	D	S	I	A	P	V	A	T	H	S	I	P	F	Y	N	S	L	T	N	A	R	
K	A	A	I	P	L	T	P	W	H	L	N	K	N	W	S	S	V	R	V	P	T	L	I	I	G	A	D	L	D	T	I	A	P	V	L	T	H	A	R	P	F	Y	N	S	L	P	T	S	I	
K	A	A	I	V	P	L	T	P	W	H	T	D	K	T	F	N	-	T	S	V	P	V	L	I	V	G	A	E	N	D	T	V	A	P	V	S	Q	H	A	I	P	F	Y	Q	N	L	P	S	T	T
Q	A	A	I	P	L	T	P	W	H	T	D	K	T	W	G	S	V	R	V	P	T	L	I	I	G	A	E	N	D	T	I	A	S	V	R	S	H	S	E	P	F	Y	N	S	L	P	G	S	L	
K	A	A	I	P	L	T	P	W	H	L	N	K	S	W	R	D	I	T	V	P	T	L	I	I	G	A	E	Y	D	T	I	A	S	V	T	L	H	S	K	P	F	Y	N	S	L	P	S	P	T	
K	A	A	I	P	L	T	P	W	H	L	N	K	S	W	R	D	I	T	V	P	T	L	I	I	G	A	D	L	D	T	I	A	P	V	S	H	S	E	P	F	Y	N	S	L	P	S	S	I		
K	A	A	I	P	L	T	P	W	H	L	N	K	S	W	R	D	I	T	V	P	T	L	I	I	G	A	D	L	D	T	I	A	P	V	L	T	H	A	R	P	F	Y	N	S	L	P	T	S	I	
K	A	A	I	P	L	T	P	W	H	L	N	K	S	W	R	D	I	T	V	P	T	L	I	I	G	A	E	Y	D	T	I	A	S	V	T	L	H	S	K	P	F	Y	N	S	L	P	S	P	T	
S	A	A	I	P	L	A	P	W	H	S	G	N	P	F	D	Q	I	E	T	P	T	M	I	I	A	C	E	S	D	S	T	A	P	V	G	S	H	A	S	P	F	Y	N	R	I	P	A	G	S	T
K	A	A	I	P	L	A	P	W	Y	A	G	N	P	F	D	E	I	T	T	P	T	L	I	I	A	C	E	L	D	V	V	A	P	V	A	Q	H	A	S	P	F	Y	R	E	I	P	G	S	T	
K	A	I	L	P	L	T	P	W	N	L	S	T	N	F	S	G	V	O	V	P	T	L	I	V	G	A	D	G	D	A	V	A	P	V	A	S	H	A	R	P	F	Y	A	S	L	P	S	T	V	
K	A	A	I	P	L	A	P	W	D	S	S	T	Q	F	R	R	S	I	T	A	P	T	L	I	F	A	C	E	N	D	A	I	A	P	V	R	S	H	A	D	P	F	Y	N	A	I	P	S	S	T
30	K	A	I	I	P	Q	A	P	W	H	T	G	S	T	F	R	R	L	T	T	P	T	M	I	I	A	C	E	A	D	V	V	A	P	V	G	Q	H	A	S	P	F	Y	N	A	I	P	D	S	T
S	A	A	I	P	I	A	P	W	N	Q	G	N	R	F	D	Q	I	E	T	P	T	L	V	I	A	C	E	N	D	V	V	A	S	V	N	S	H	A	S	P	F	Y	N	R	I	P	S	T	T	
K	A	A	I	P	L	A	P	W	D	T	T	-	S	F	R	N	V	K	A	P	T	L	I	F	A	C	E	S	D	V	I	A	P	V	G	Q	H	A	S	P	F	Y	N	R	I	P	S	N	I	
K	A	A	I	P	Q	A	P	W	Y	S	G	N	S	F	N	R	I	T	T	P	T	L	I	I	A	C	E	L	D	V	V	A	P	V	G	Q	H	A	S	P	F	Y	N	R	I	P	S	S	T	
K	A	I	I	P	Q	A	P	W	Y	S	G	N	S	F	D	R	I	T	T	P	T	M	I	I	A	C	E	S	D	A	I	A	P	V	G	Q	H	A	S	P	F	Y	N	D	I	P	N	S	T	
K	A	A	I	P	F	A	P	W	N	T	V	T	N	W	S	A	V	T	V	P	T	L	I	M	A	C	O	A	D	A	I	A	P	V	A	S	H	A	V	F	Y	N	M	-	S	A	N			
K	A	A	V	P	F	A	P	W	H	N	T	A	N	F	S	G	V	O	V	P	T	L	V	I	A	C	E	N	D	T	V	A	P	I	S	R	H	A	S	S	F	Y	N	S	L	P	S	S	L	
K	A	S	I	P	F	A	P	W	N	S	T	E	N	F	S	S	V	T	V	P	T	L	I	I	A	C	O	D	D	S	I	A	P	V	Y	S	H	A	S	P	F	Y	N	S	M	T	R	N	-	
R	A	A	A	P	Q	A	P	W	A	Q	-	E	S	F	S	S	M	T	V	P	T	L	I	V	S	C	E	N	D	S	I	A	P	N	S	S	H	S	A	P	F	Y	N	Q	M	T	R	N	-	
K	A	A	A	P	Q	A	P	W	N	S	S	T	N	F	S	S	M	T	V	P	T	L	I	I	A	C	E	S	D	S	T	A	P	T	R	S	H	A	E	P	F	Y	N	S	M	S	R	N	-	
K	A	A	A	P	Q	A	P	W	N	N	S	T	N	F	S	S	V	T	V	P	T	L	I	L	A	C	E	S	D	S	T	A	P	V	N	Q	H	A	S	P	F	Y	N	S	M	S	R	N	-	
K	A	V	V	P	L	A	P	W	Y	V	G	G	L	E	Q	S	T	V	P	T	M	I	F	G	A	D	N	D	F	I	A	P	V	A	S	N	A	R	P	F	Y	Q	S	L	T	K	V	P	I	
K	A	A	I	P	L	A	P	W	D	T	T	Y	A	S	R	S	O	A	A	P	T	L	I	F	A	C	E	S	D	V	I	A	P	V	L	O	H	A	S	P	F	Y	N	S	L	P	S	P	I	
K	A	A	I	P	L	T	P	W	H	L	D	K	T	W	S	E	V	D	V	P	T	M	I	I	G	A	E	N	D	T	V	A	S	V	S	H	S	I	P	F	Y	N	S	L	S	G	A	P		
Q	A	A	V	P	L	A	P	W	N	L	D	K	T	W	S	G	V	T	V	P	T	L	I	V	G	A	E	S	D	S	V	A	P	V	A	S	H	S	V	P	F	Y	N	S	L	P	A	S	S	A
Q	A	A	V	P	L	A	P	W	N	L	D	K	T	W	S	G	V	T	V	P	T	L	I	V	G	A	E	S	D	S	V	A	P	V	A	S	H	S	V	P	F	Y	N	S	L	P	A	S	S	A
K	A	S	I	P	L	T	P	W	N	L	K	N	W	R	N	N	R	V	P	T	L	I	V	G	A	E	N	D	S	I	A	S	V	R	T	H	A	E	P	F	Y	E	S	I	P	S	T	T		
Q	A	A	V	P	L	A	P	W	N	L	D	K	T	W	S	S	L	R	V	P	T	L	I	V	G	A	E	S	D	S	V	A	P	V	S	H	S	V	P	F	Y	N	S	L	P	A	G	A	T	
K	A	A	I	P	L	A	P	W	N	L	D	K	T	W	S	P	E	V	S	T	P	T	L	V	V	G	A	Q	D	D	S	I	A	S	V	T	T	H	A	E	P	F	Y	E	S	L	P	G	A	T
R	A	A	I	P	L	T	P	W	H	L	Q	K	N	W	S	D	V	E	V	P	T	M	I	I	G	A	E	N	D	T	V	A	S	V	R	T	H	S	I	P	F	Y	E	S	L	D	E	D	L	
10	Q	A	A	I	P	L	T	G	W	H	T	T	K	N	W	S	T	V	R	V	P	T	M	V	I	G	A	E	N	D	S	I	A	P	V	A	Q	H	S	E	P	F	Y	T	S	L	P	S	T	L
K	A	I	I	P	M	T	P	W	N	L	D	K	T	W	S	G	D	V	E	V	P	M	L	D	F	G	A	E	Y	D	T	I	A	S	P	T	T	H	A	K	P	F	Y	N	-	H	G	-		
Q	A	A	I	P	L	T	G	W	H	T	V	K	N	W	S	S	V	R	V	P	T	L	V	V	G	A	E	N	D	T	V	A	P	V	S	S	H	S	E	P	F	Y	T	S	L	P	A	T	L	
R	A	A	I	P	L	T	P	W	H	T	O	K	N	W	S	S	V	O	T	P	T	L	V	V	G	A	Q	N	D	V	T	A	S	V	R	S	H	A	I	P	M	Y	E	S	L	P	S	S	L	
K	A	A	I	P	L	A	P	W	N	L	T	K	S	W	T	R	O	R	V	P	T	V	I	V	G	S	S	D	T	I	A	P	V	S	H	S	E	P	F	Y	T	S	L	-	G	S	P			
Q	A	A	V	P	L	T	P	W	H	L	D	K	T	W	N	E	V	R	V	P	T	M	I	I	G	A	E	A	D	T	T	A	S	V	G	S	H	S	E	R	F	Y	A	S	L	T	N	A	S	
K	A	A	V	A	L	Q	P	W	H	T	D	K	T	W	P	E	V	R	V	P	T	M	I	I	G	A	E	D	D	T	I	A	P	V	A	S	H	S	I	P	F	Y	T	S	A	T	G	A	R	
R	A	A	G	V	P	L	A	P	W	H	T	T	S	W	P	H	V	T	N	P	W	M	I	L	G	G	O	N	D	G	I	A	P	V	S	H	A	I	P	M	Y	T	G	V	A	S	G	-		
K	A	A	I	P	L	A	P	W	N	L	D	K	T	W	P	E	V	R	T	P	T	L	I	I	G	G	E	L	D	A	V	A	P	V	A	T	H	S	I	P	F	Y	N	S	L	N	A	P		
Q	A	A	V	P	I	A	P	W	N	L	T	K	S	W	S	G	L	R	V	P	T	M	I	V	G	G	E	S	D	S	I	A	P	V	S	H	S	I	P	F	Y	T	S	I	P	A	S	A		
Q	A	A	V	P	L	A	P	W	H	T	T	K	N	W	S	G	V	R	V	P	T	M	I	I	G	G	E	S	D	S	V	A	P	V	S	T	H	S	E	L	F	Y	A	S	L	P	A	S	S	
K	A	A	I	P	L	A	P	W	N	T	D	K	T	W	P	E	V	R	T	P	T	L	I	I	G	G	E	S	D																					



200

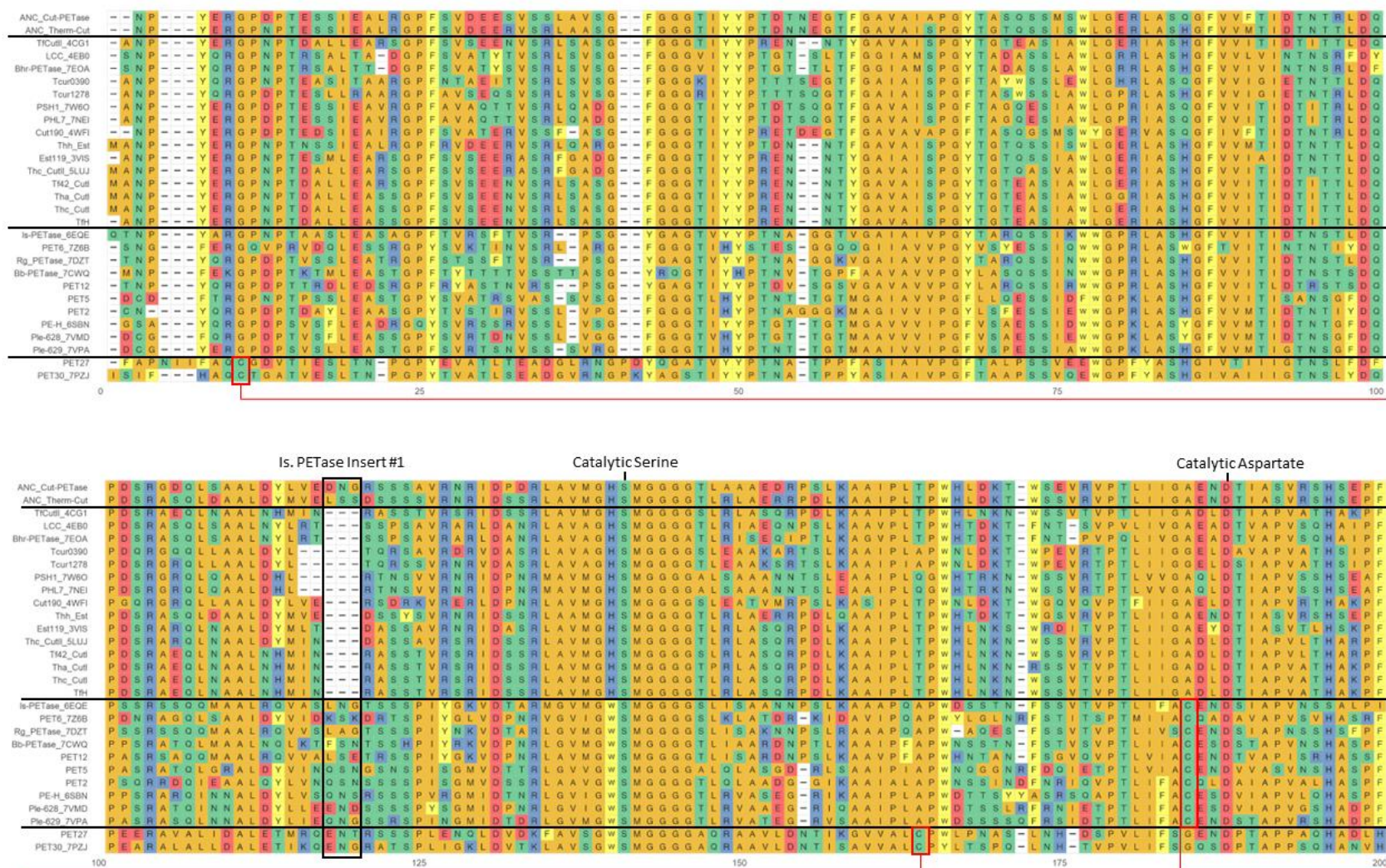
220

240

260

Figure A2 Multiple Sequence Alignment for Cut-Therm. Alignment above contains sequence homologs (based on Tf. Cut II 30-95% identity, coverage > 75%, E-Val < 1E-20) with the following functional labels: cutinase (24), PETase (1), thermo (25), lipase (25), polyethylene terephthalate (1), polyester (1), and closest PETase homologs (25). This alignment was created using MAFFT. In this depiction, residues matching consensus (shown in the top row), are colored by their side-chain properties. Important sites such as catalytic residues, and insert loops are annotated.

D. Alignment of Known PET Hydrolases



E. Phylogenetic Trees

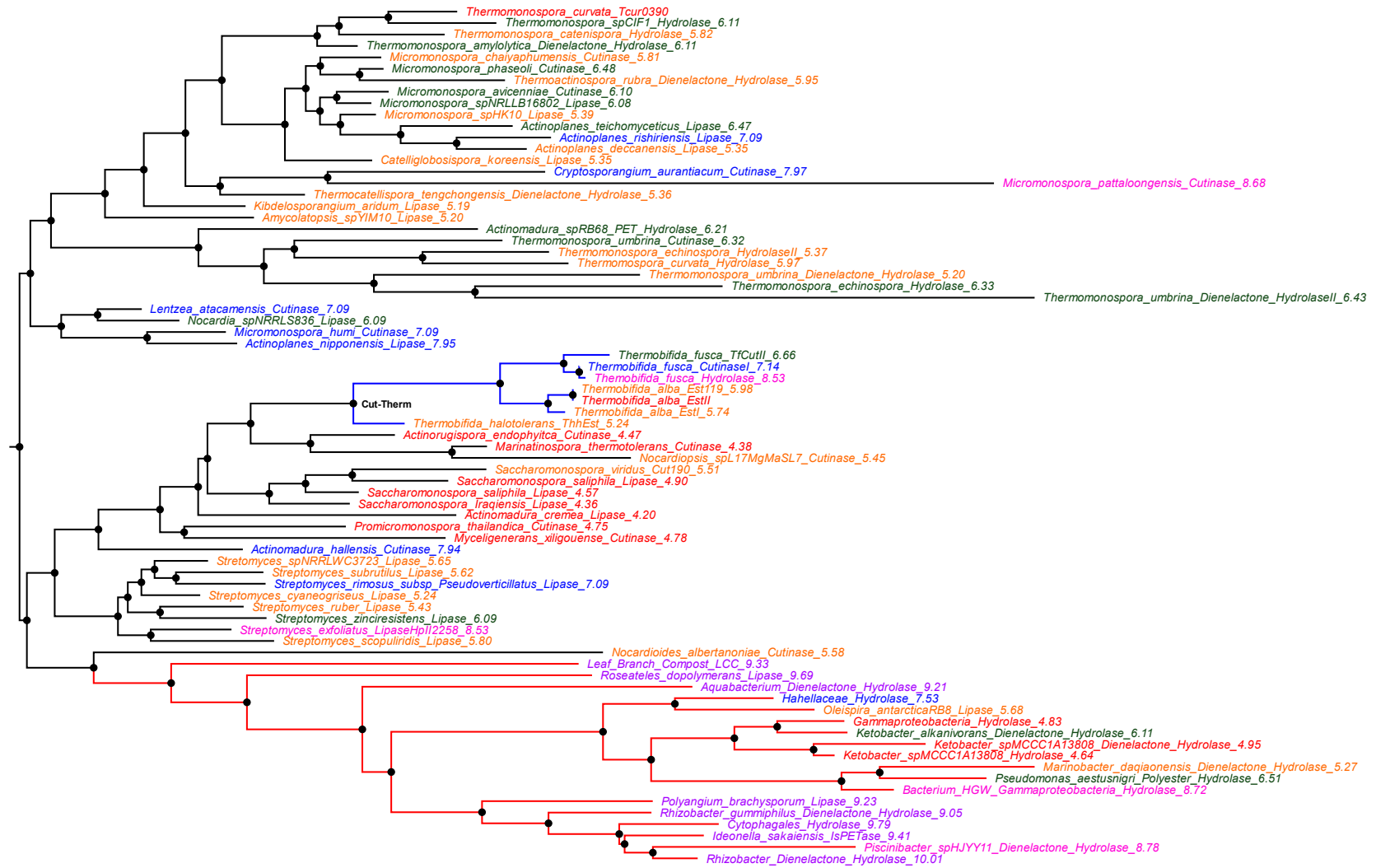


Figure A4 pI Representation of Cut-Therm Phylogenetic Tree. A representation of the Cut-Therm phylogenetic tree colored by pI of sequences: < 5 = red, 5-6 = orange, 6-8 = blue, 8-9 = pink, and >9 = purple. This was completed to determine where the low pI of ancestors (4.95 for ANC Cut-Therm) was originating from. Indeed, the majority of sequences surrounding the ANC Therm Cut node (red highlight), have a pI between 4-6. This tree was built IQTree, with evolutionary model LG+F+I+G4; the figure was created using Python (for colouring) and FigTree.

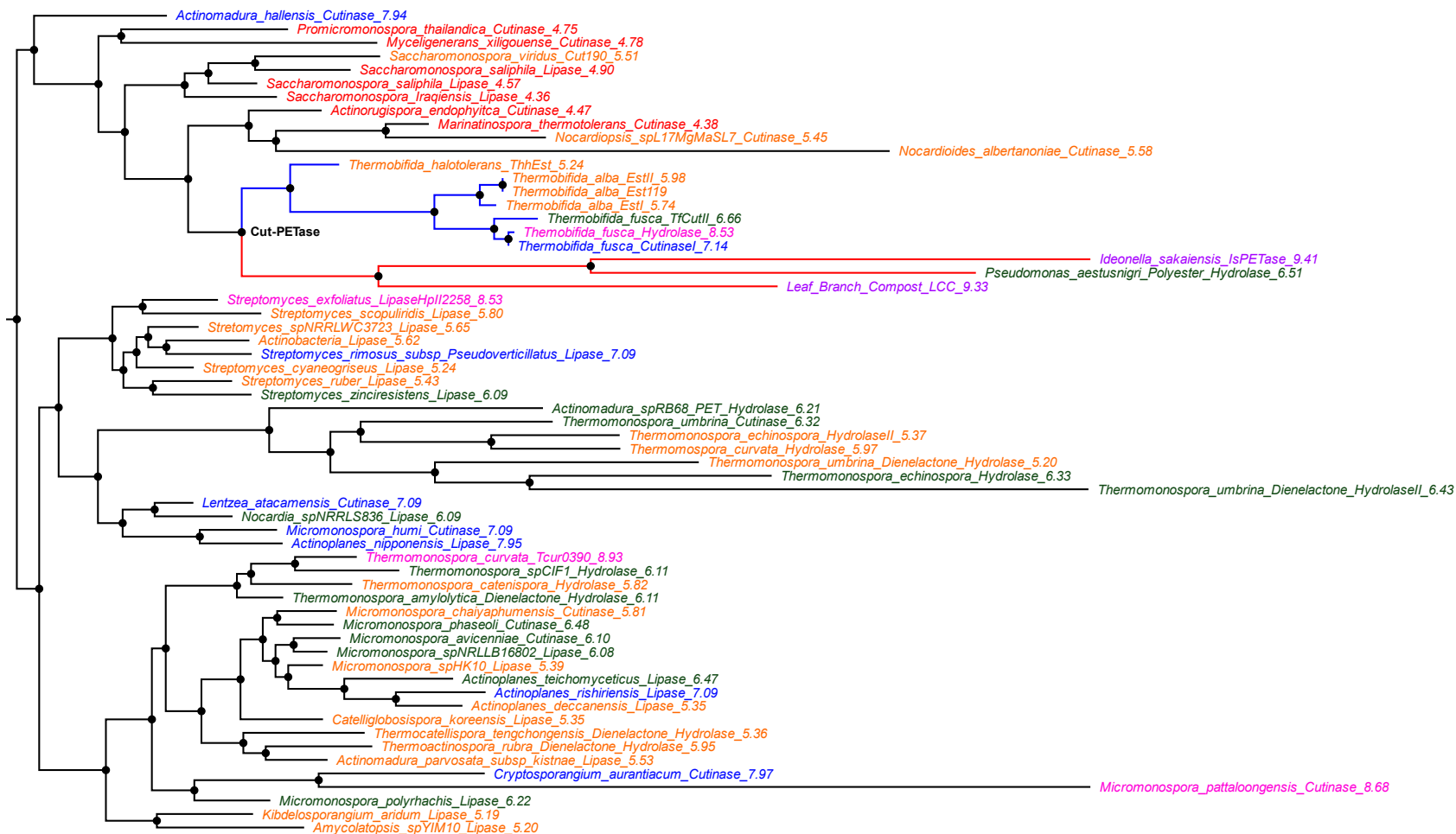


Figure A5 pI Representation of Cut-PETase Phylogenetic Tree. A representation of the Cut-PETase phylogenetic tree colored by pI of sequences: < 5 = red, 5-6 = orange, 6-8 = blue, 8-9 = pink, and >9 = purple. This was completed to determine where the low pI of ancestors (4.90 for ANC Cut-PETase) was originating from. While the ANC Cut-PETase node has high pI descendants, the majority of surrounding and directly upstream sequences have low pI values (4-6). This tree was built IQTree, with evolutionary model LG+F+I+G4; the figure was created using Python (for colouring) and FigTree.

F. Mass Spectrometry

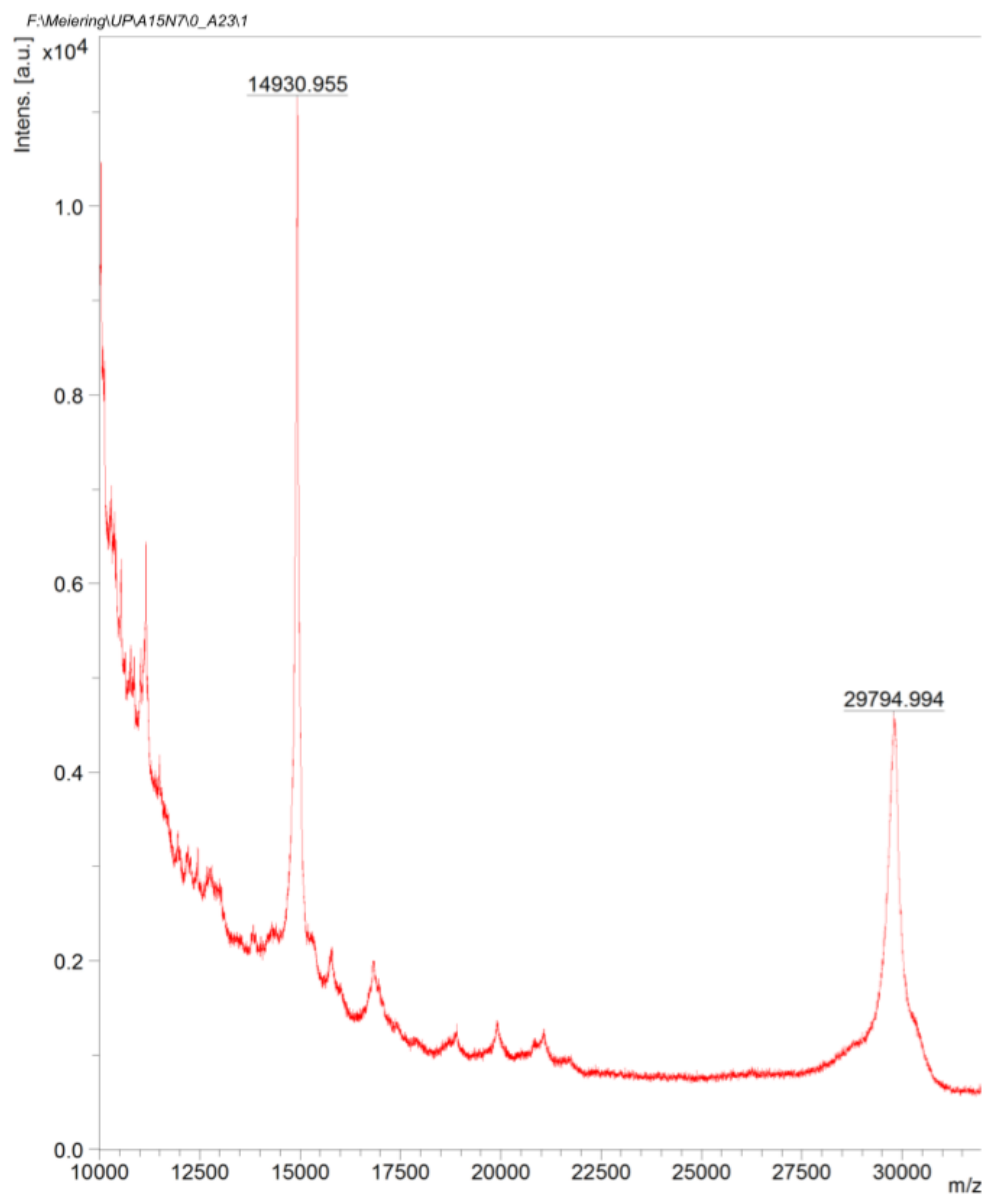


Figure A6 MALDI mass spectrometry analysis of ANC Cut-PETase. MALDI was run on ANC Cut-PETase (buffer exchanged in water), using the Bruker Autoflex Speed spectrometer. The peak associated with a m/z of 29794 is expected to be the singly charged result for the protein. The 14930.955 m/z peak was described by Valerie to be the doubly-charged peak, which has been observed for proteins before in linear-mode MALDI (Mädler, Barylyuk, ... Zenobi, 2012). This said, the intensities of double-charged peaks in literature are typically smaller than monoisotopic peaks, while the one seen here is substantially larger. The cause of this remains unclear. However, given the lack of a correlating band in SDS-page for a smaller protein, and the peaks positioning at the half m/z for the singly charged peak it is probable that this is indeed the doubly charged peak.

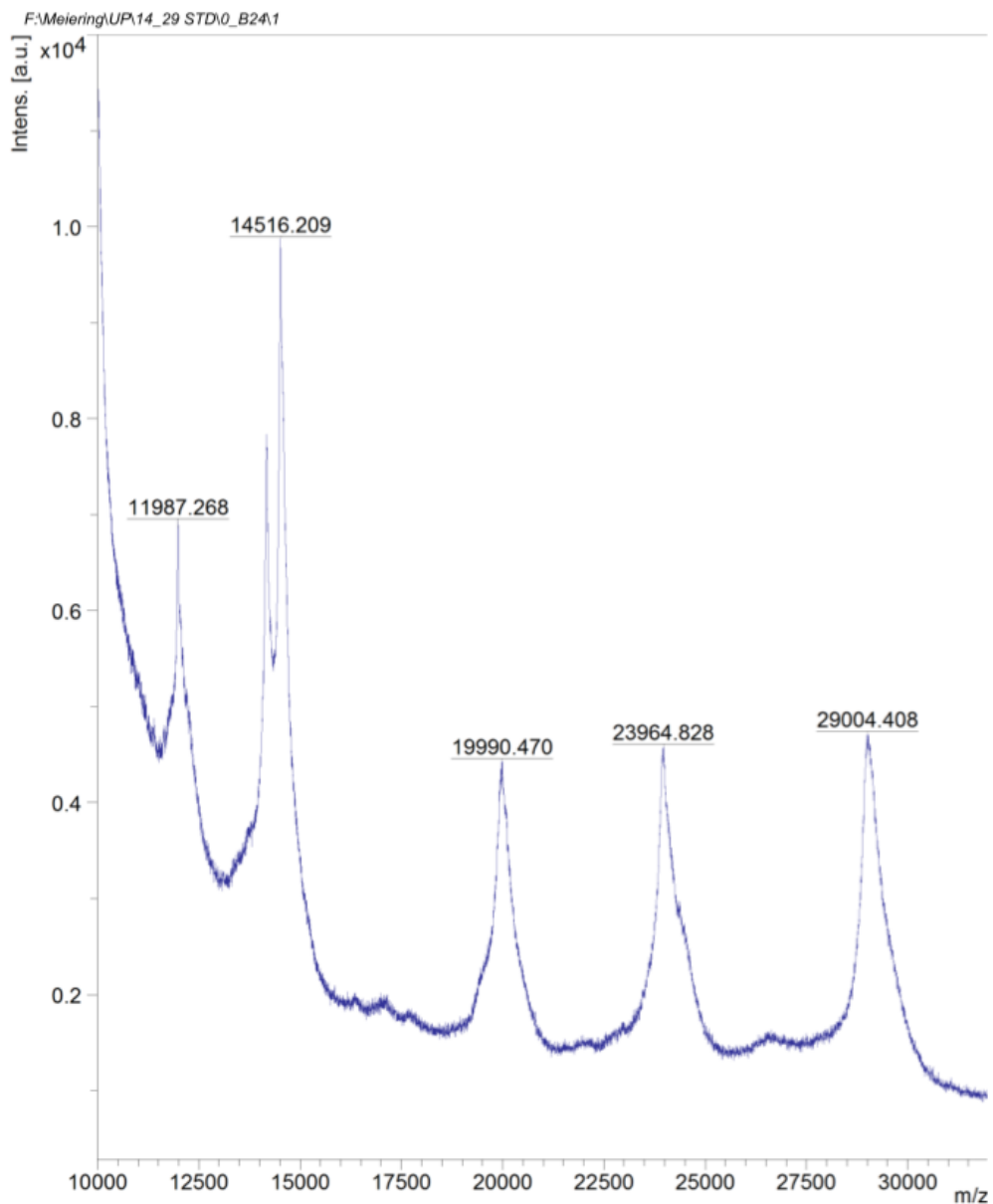


Figure A7 MALDI mass spectrometry of protein standards. Protein standards with molecular weights of 14.177, 19.977, 23.982, and 29.024 kDa were run using linear-mode MALDI on the Bruker Autoflex Speed spectrometer. The 11987.26 m/z peak and the small peak immediately before the 14516.209 m/z peak are likely the doubly charged peaks for the 23964.828 and 29004.408 peaks respectively.

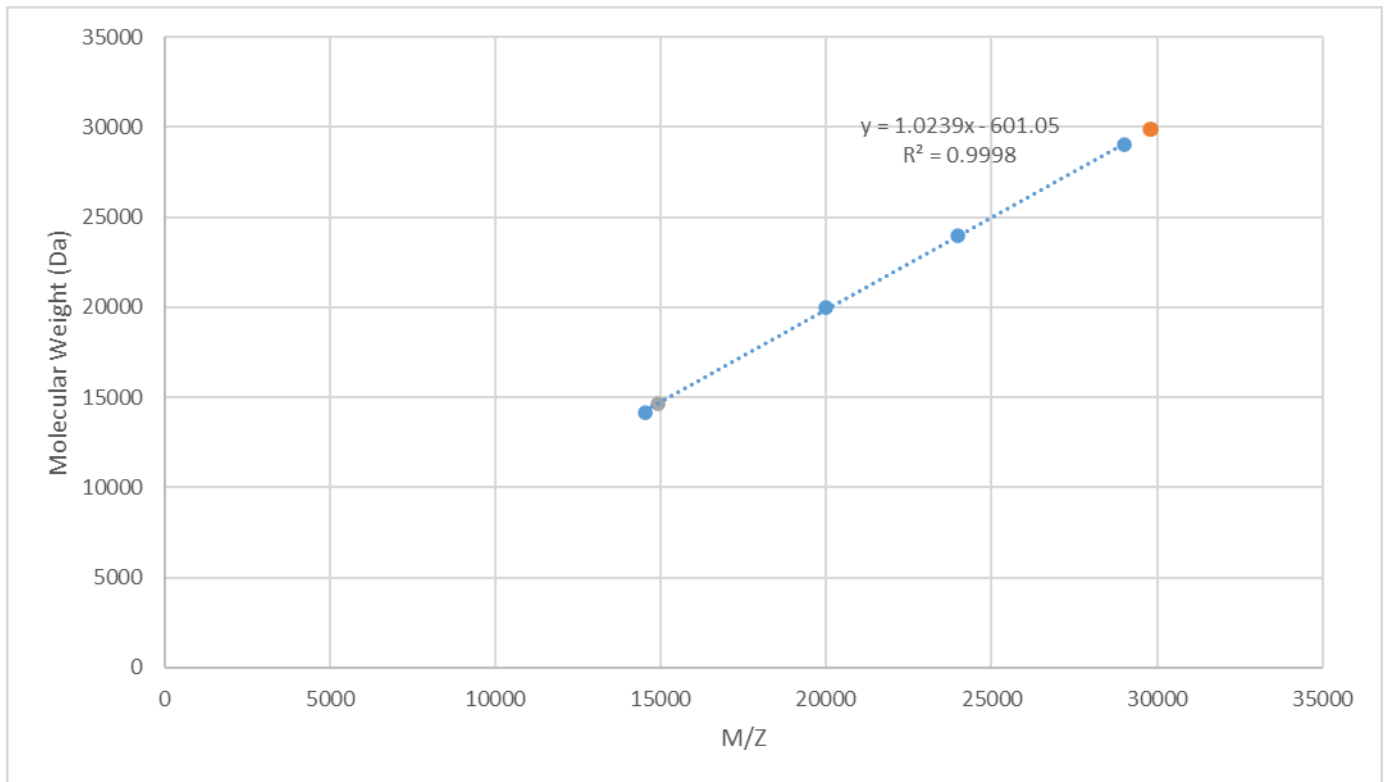


Figure A8 Standard curve relating MALDI m/z to molecular weight (Da). Standard curve (blue) was generated from MALDI data on proteins with molecular weights of 14.177, 19.977, 23.982, and 29.024 kDa. Standard data fit well into a linear model ($y = 1.0239x - 601.05$), with a R^2 value of 0.9998. The orange dot and grey dots relate to the singly and doubly charged ANC Cut-PETase m/z values, and correspond to molecular weights of 29.91 and 29.99 kDa respectively. The expected molecular weight for ANC Cut-PETase is 29.89 kDa, which is within 20 Da of the predicted singly charged mass.

G. Location of Tryptophan Residues in Tf. Cut II and Ancestors

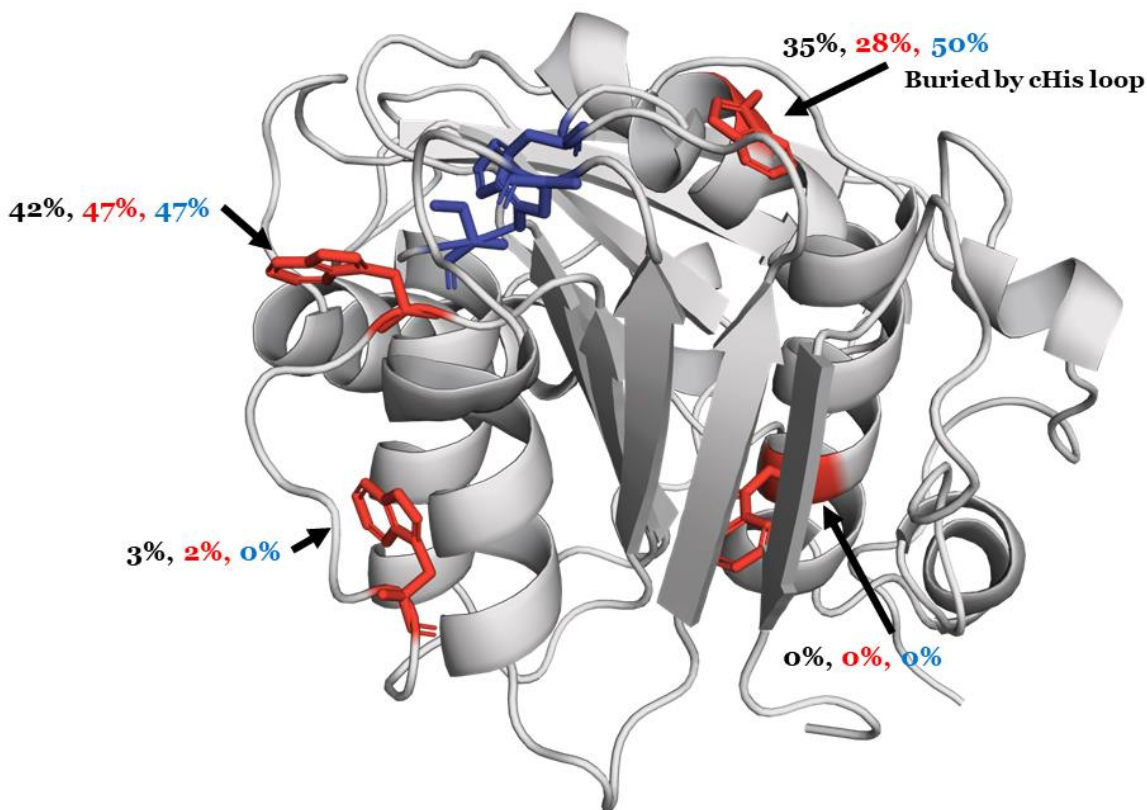


Figure A9 Location of Trp Residues on Tf. Cut II and Ancestors Cartoon representation of Tf. Cut II above shows shared location of four Trp residues (red sticks): probing environment of Trp residues is possible during fluorescence measurements. Between Tf. Cut II and ancestors, the locations on the protein for these Trp residues is conserved. CAMSOL-SASA (defined as the exposure of a residue in the enzyme compared to the same residue in a G-X-G peptide (Sormanni, Aprile, and Vendruscolo, 2015)) for Tf. Cut II, ANC Cut-PETase, and ANC Cut-Therm are labelled in black, red, and blue respectively. Interestingly, $\frac{3}{4}$ Trp residues have similar solvent exposure predictions. The remaining Trp, has SASA values ranging from 28, 35, and 50%, which suggests structural variation in the loop responsible for burial of this Trp site (being the active site cHis loop) may be possible. Active site residues are shown as blue sticks.

H. Temperature denaturation of Tf. Cut II

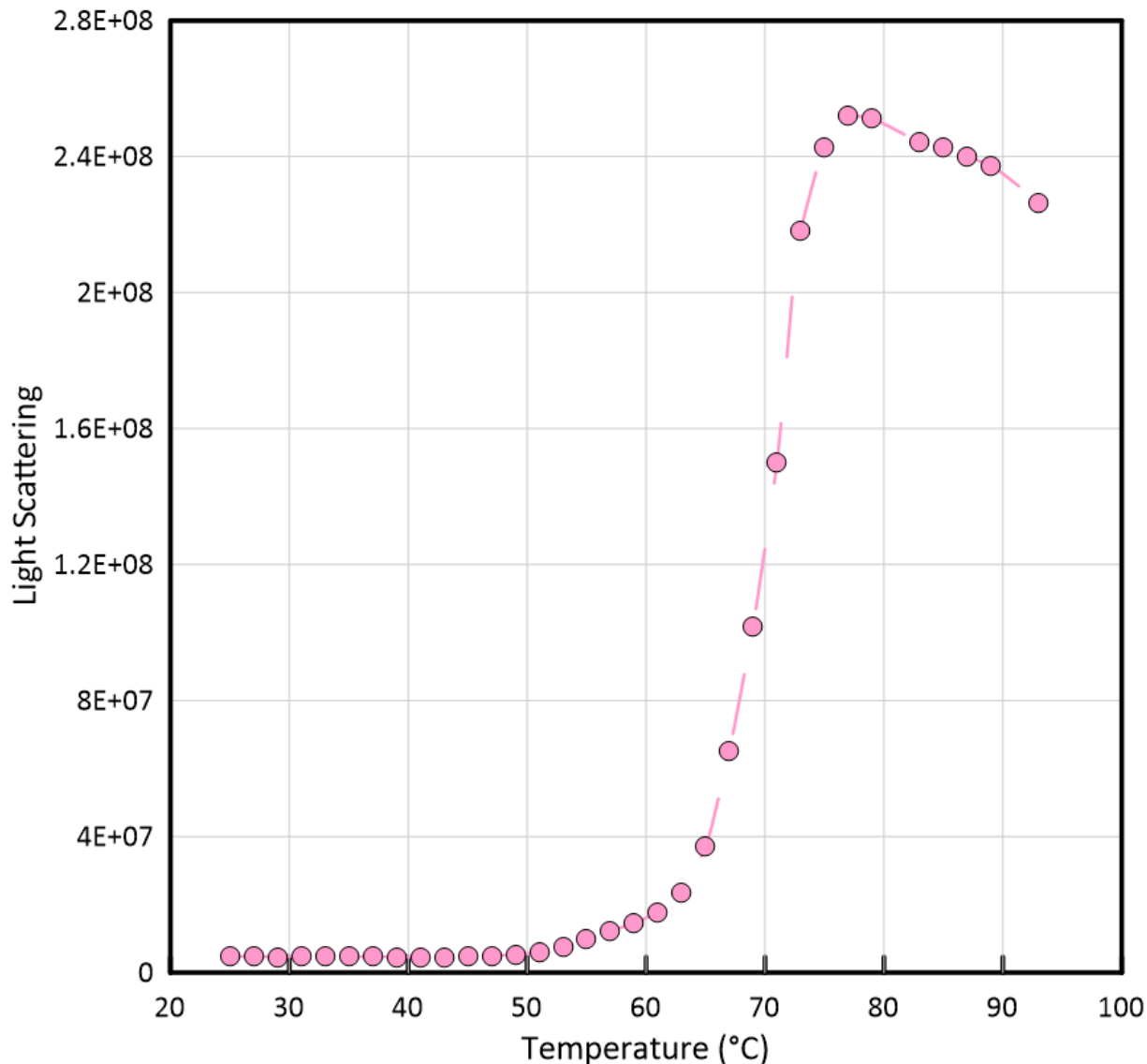


Figure A10 Increase in Tf. Cut II light scattering with temperature may correlate to T_m . Here, light scattering measurements were performed using a Horiba fluorimeter (model FL3-11), in a 1-cm pathlength cuvette, 450 em, 450 ex. A sample of Tf. Cut II ($3 \mu\text{M}$) was placed in the cuvette, and progressively heated at an approximate rate of $1 \text{ }^\circ\text{C}/\text{minute}$ (controlled by a Peltier module). Light scattering measurements were collected every $2 \text{ }^\circ\text{C}$ for 30 seconds, for a range of $25\text{-}95 \text{ }^\circ\text{C}$. The results are expected to be correlated with aggregation, which can significantly increase following/during denaturation (due to exposure of more hydrophobic, buried residues). This analysis reveals an estimated, apparent denaturation temperature around $70 \text{ }^\circ\text{C}$, which correlates to apparent T_m estimations in literature (Roth et al. 2014).

I. Iodoacetamide labelling and gel electrophoresis

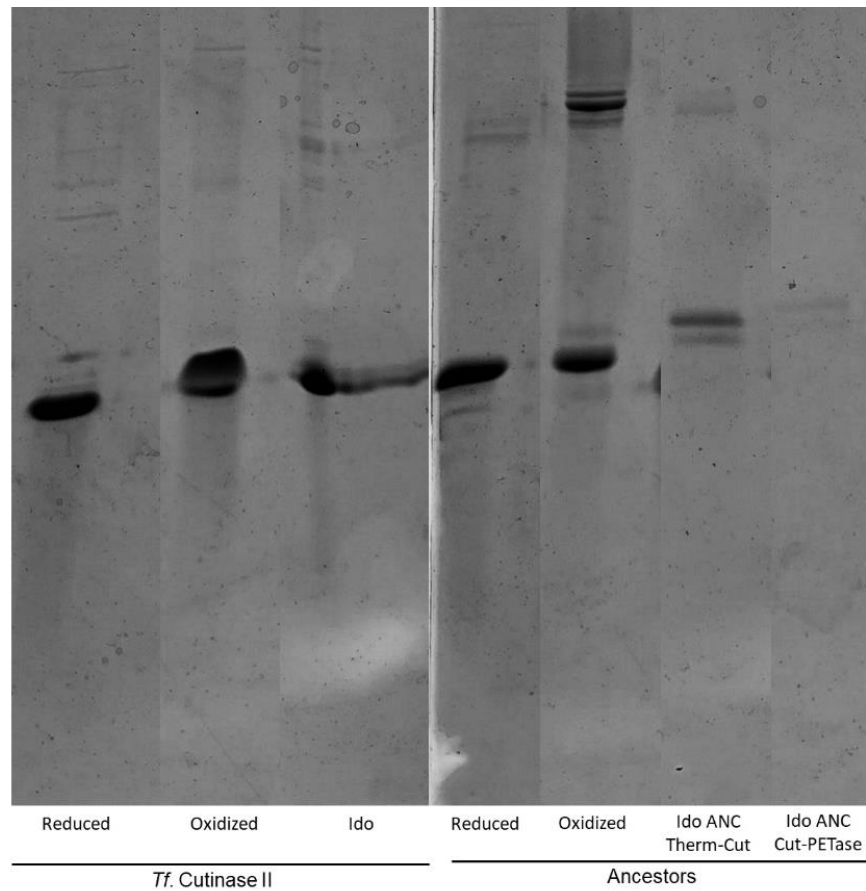


Figure A11 Iodo-acetamide labelling of post-chemical denaturation samples reveal ancestor disulfides were mainly reduced. Post chemical denaturation samples for ancestors and Tf. Cut II were analyzed by iodo-acetamide labelling. Essentially, in this analysis exposed thiols on Cys residues can react with iodo-acetamide to form labelled proteins. These will run slightly heavier on an SDS-page gel compared to their non-labelled counterparts, and thus can be used to assess the state of disulfide bonds in a protein. Here, a majority of the tested ancestor samples (right lanes, labelled “IDO”), run heavier than their oxidized/reduced unlabelled counterparts, suggesting that they were indeed reduced during analysis (note that the ANC Cut-PETase sample is quite faint). On the other hand, the left lanes, reveal that the wildtype was likely unreduced due to similar movements between analyzed (IDO) and unlabelled (reduced, oxidized) samples. A 15% resolving gel was used for this analysis.

J. Structural comparison of ancestor homology models to Tf. Cut II and Is. PETase

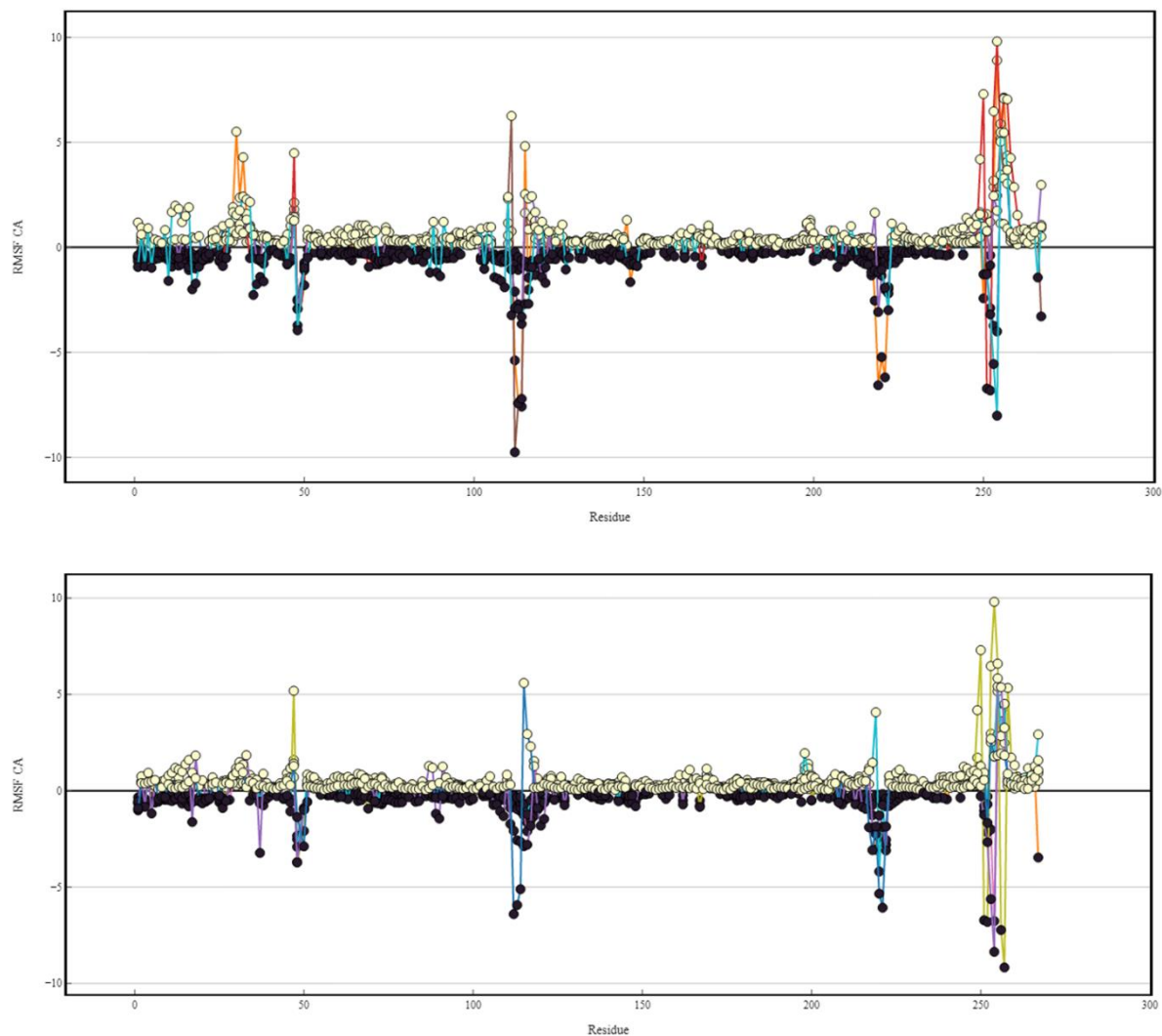


Figure A12 Deviation of predicted ancestor homology models from Tf. Cut II and Is. PETase structures. Top scored homology models from SWISS-Prot, Alpha Fold, I-Tasser, Phyre, Robetta CM, and Robetta TTA were aligned to structures of Tf. Cut II and Is. PETase separately in Pymol. Then, distances between ancestor and Tf. Cut II/Is. PETase Ca were obtained. These values were graphed in python such that distances for residues closer to Tf. Cut II Cas were plotted positively (yellow), and distances for residues closer to Is. PETase were plotted negatively (black). This analysis was used to (1) find areas where models disagreed, and (2) to determine their likeness to existing enzymes. Here, results for ANC Cut-PETase are **in the top graph**, while those for ANC Cut-Therm are **on the bottom**. All sites with a deviation greater than 2 Å from a wild-type protein were looked into individually.

K. Summary for Models

Table A2 MolProbity, Fold-X, and Rosetta Scores for Ancestor Homology Models The below tables summarize homology model results for MolProbity Ramachandran scores (first section), and energy scores (FoldX, Rosetta). In all cases, when applicable, particularly poor scores were coloured red, while comparably better scores were coloured green. This analysis revealed that I-Tasser scored quite poorly overall. Likewise, Phyre2 has more Ramachandran outliers, which is considered undesirable. The remaining four models have similar MolProbity results. Intriguingly, Alpha Fold 2 models have a higher number of bad bonds (referring to outliers in distances between connected atoms). It is unclear how significant this value is (especially almost all of these ‘band bonds’ are not severe), and whether it is sufficient to rationalize against use of Alpha Fold2 models based on this premise. Ultimately, Alpha Fold2 models were chosen, due to its superior performance over Robetta in CASP14 (Pereira, Simpkin, ... Lupas, 2021).

ANC Cut- PETase	Tf. Cut II	Alpha Fold2	Swiss-Prot	Robetta CM	Robetta TTA	Phyre2	I-Tasser
Poor Rotamers	4	0	2	0	0	0	21
Favoured Rotamers	206	218	214	221	220	221	177
Rama Outliers	1	1	1	0	1	6	18
Rama Favoured	248	258	253	225	253	248	215
Rama Z- Score	1.14	-0.33	0.05	0.08	1.29	-0.41	-4.33
Cb Deviations	2	0	7	0	0	0	9
Bad Bonds	14	57	2	3	4	16	0
Bad Angles	45	7	28	7	6	7	31
Cis-Prolines	2	1	2	0	1	1	2
Twisted Peptides	0	0	0	1	0	0	9
FoldX Energy	180.87	156.87	86.44	82.81	70.76	328.60	178.19
Rosetta Energy	-481	537	-187	-834	-830	2665	540

ANC Cut-Therm	Tf. Cut II	Alpha Fold2	Swiss-Prot	Robetta CM	Robetta TTA	Phyre2	I-Tasser
Poor Rotamers	4	0	2	0	0	0	27
Favoured Rotamers	206	222	219	224	225	225	176
Rama Outliers	1	1	0	1	0	5	15
Rama Favoured	248	258	250	257	256	249	225
Rama Z-Score	1.14	-0.3	-0.4	0.5	0.39	-0.38	-3.99
Cb Deviations	2	0	2	0	0	0	8
Bad Bonds	14	54	1	6	1	18	0
Bad Angles	45	3	11	11	1	6	37
Cis-Prolines	2	1	2	0	1	1	0
Twisted Peptides	0	0	0	0	0	0	10
FoldX Energy	180.87	180.68	90.53	72.83	62.35	371.86	189.89
Rosetta Energy	-481	1778	-266	-839	-868	3032	717

L. Key data related to Ancestor Molecular Dynamics

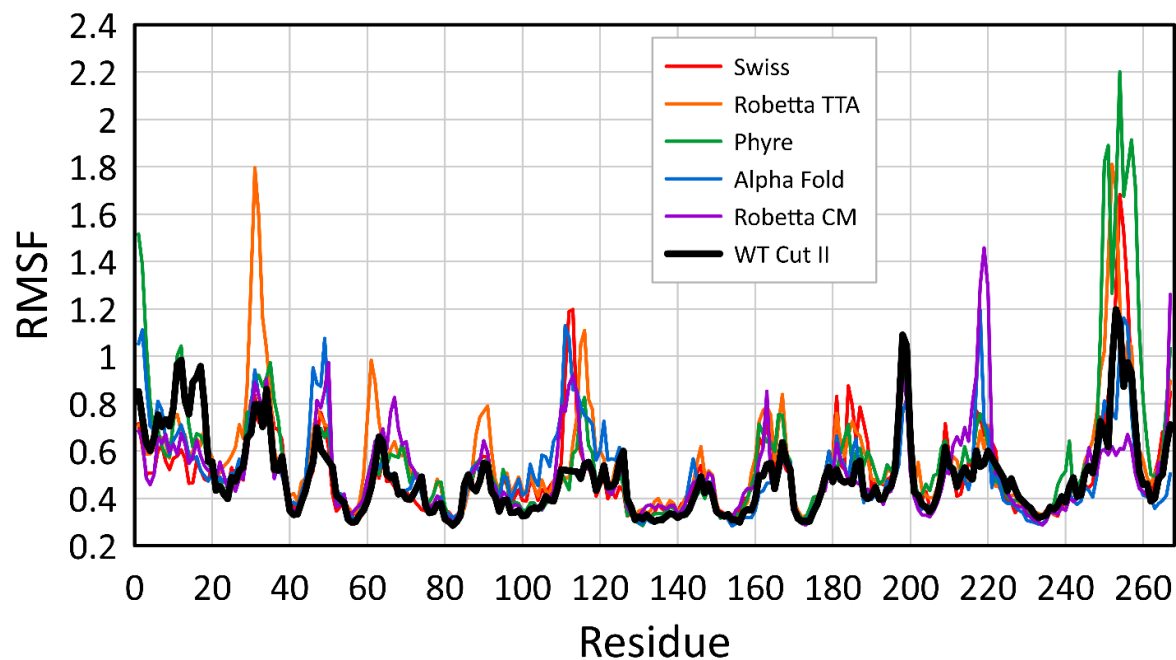
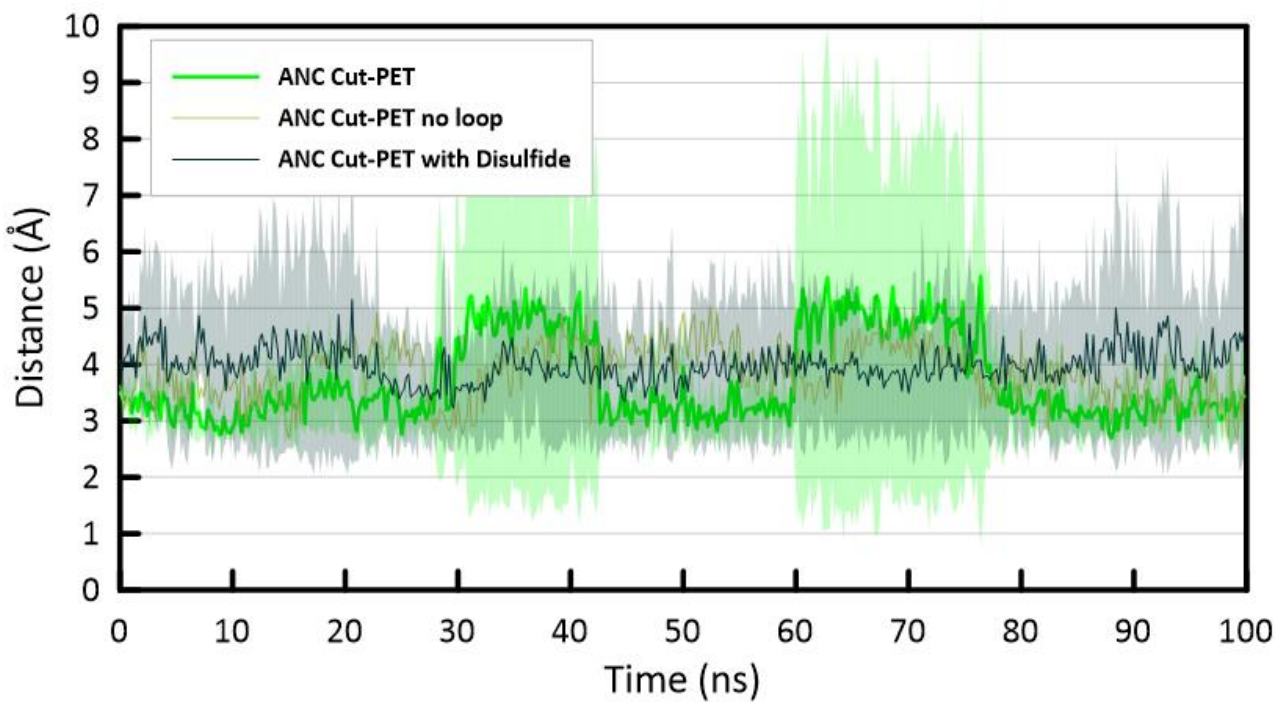
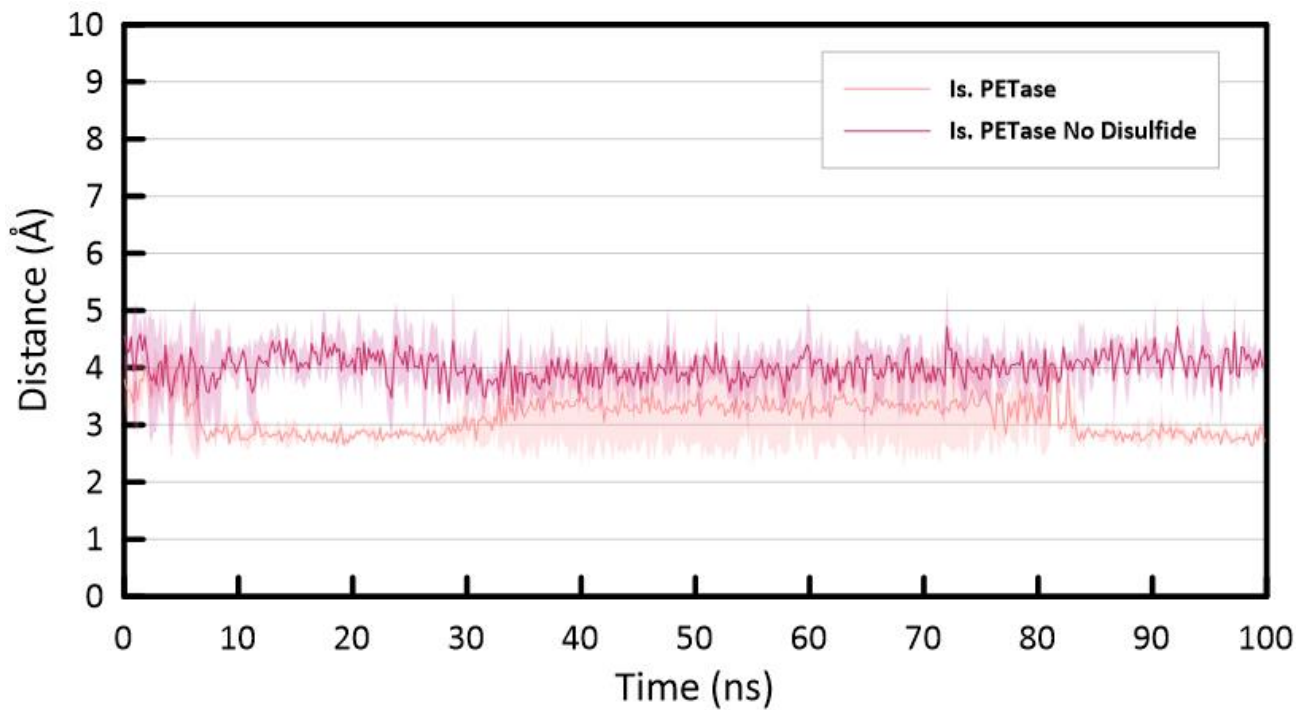


Figure A13 RMSF for all top-scoring homology models for ANC Cut-PETase from several programs. Results above reveal similar RMSF patterns for all models, with some models having higher RMSF at some regions (i.e., Robetta TTA at residues 20-30, Phyre C-terminal). As a caveat, these simulations were run only once (instead of in triplicate as for simulations used for analysis), and were only run for 30 ns. Further, at this time, equilibrations for simulations included fewer steps. However, most models agree on sites with high vs. low RMSF, suggesting that these may be sites with higher fluctuations (especially when compared to Tf. Cut II (black)).



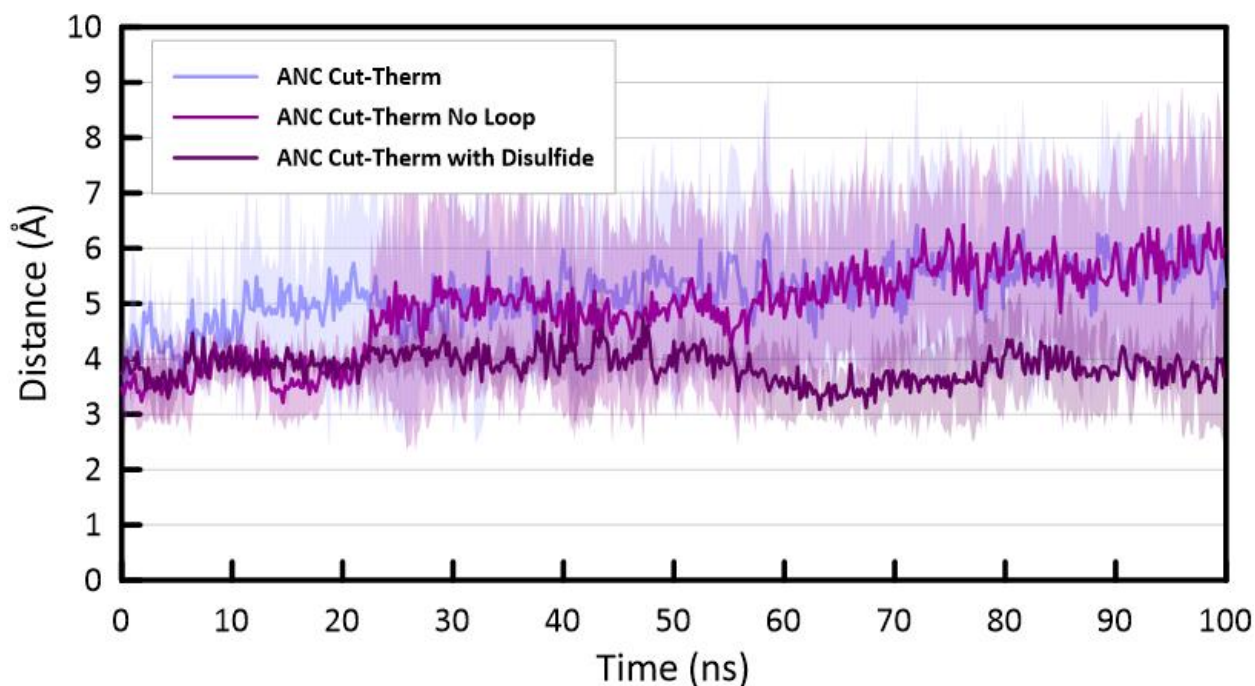


Figure A14 Changes to active site loop may affect catalytic triad for ANC Cut-Therm more than ANC Cut-PETase. Distance between cSer and cHis in 100 ns simulations are used here to assess the stability of the active site catalytic triad. All simulations for this analysis were run in triplicate (except for ANC Cut-PET no loop, which was only performed once), using the method described in the Materials and Methods. **(Top)** In *Is. PETase* simulations, the distance is typically stabilized at an average of 3.2 Å, and reaches a minimum of 2.5 Å. Upon removal of the active site disulfide the average distance is increased to 4.01, but remains quite consistent. Alanine mutations to this disulfide has resulted in inactive variants, which may be due to disconnection of the catalytic triad (Fecker, Galaz-Davison, ... Ramírez-Sarmiento, 2018). Ancestor results show distances similar to or beyond the no disulfide *Is. PETase*, and show greater amounts of deviation. This may suggest that the health of the triad is not sufficient for catalysis. **(Middle)** Distances between cSer and cHis in ANC Cut-PETase for ancestor (light green, average = 3.7 Å), ancestor without loop inserts (grass green, average = 3.9 Å), and with the additional active site disulfide added (dark green, average = 3.9 Å). Changes to average distance upon loop permutations are small, however, consistency of distance may improve upon removal of loops or addition of disulfide. The minimum distances for ANC Cut-PETase and ANC Cut-PETase with no loop reaches 2.6 Å, while the lowest distance reached with disulfide inclusion was 3 Å. **(Bottom)** Distances between cSer and cHis in ANC Cut-Therm for ancestor (light purple, average = 5.2), ancestor without loop inserts (pink, average = 4.9), and ancestor with disulfide (dark purple, average = 3.8). These results reveal that the addition of the active site disulfide in ANC Cut-Therm can improve the average distance as well as the consistency of this distance over three simulations, but still remains higher than distances seen in wild-type *Is. PETase*. Notably the minimum distance seen between cSer and cHis in any ANC Cut-Therm model here, are 2.9-3.0 Å.

M.ANC Cut-PETase Extended Residue Data

Table A3 Extended residue data for ANC Cut-PETase. The below table represents data collected for ANC Cut-PETase from several techniques at a residue level. **[Res #]** Refers to ANC residue #. **[dSSP]** Refers to secondary structure of ANC, where B (yellow) is beta sheet, and H (blue) is helix. **[AA]** Refers to ANC amino acid. **[Thh Est Mut]** The sequence for Thh_Est (closest ancestor) for comparison, where mutations are highlighted in red. **[ddG]** Predicted stability upon mutation of Thh Est to ancestor residue, where larger numbers are considered destabilizing (>1, red), and negative numbers are considered stabilizing (<-1, green). **[INSEQs, GAPSEQs]** Refers to known PET hydrolase INSEQ and GAPSEQ consensus residues. Changes from ancestor are colored in red. **[PROSS]** Mutations that are predicted to be stabilizing by PROSS (Goldenzweig, Goldsmith, ... Fleishman, 2016). **[ASR]** For sites with <50% likelihood for a given ancestral residue, alternative residues are listed in order of likelihood. **[Lit Stab, Lit Act]** Comparison of ancestor residues to known literature mutations, where a -1 refers to destabilizing or deactivating residue in the ancestor (red), and +1 refers to stabilizing or activating residue in ancestor (green). **[Camsol SASA]** Extent of residue burial predicted by Camsol, where low values suggest a buried residue (yellow) and high values suggest solvent exposure (blue) (Sormanni, Aprile, and Vendruscolo, 2015). **[RMSF 30, RMSF 45, RMSF 60, RMSF 75]** Residue fluctuation averages from molecular dynamics simulations, where higher fluctuations are coloured red. The number 30, 45, 60, and 75 refers to the temperature the simulation was run. **[NL, wD, nD]** Residue fluctuation averages from molecular dynamics simulations, where higher fluctuations are red. NL refers to no loop simulations, wD refers to simulations with the additional active site disulfide, and nD refers to simulations with a reduced terminal disulfide (no active site disulfide). **[Camsol, AGGRESCAN]** Aggregation propensity predictions by CAMSOL (structure corrected), and AGGRESCAN where larger values indicate higher aggregation propensity (red) (Conchillo-Solé, de Groot, ... Ventura, 2007; Sormanni, Aprile, and Vendruscolo, 2015). **[Pocket 0ns, Pocket 100 ns]** Residue involvement with pockets in alpha fold model (0 ns), compared to pocket involvement following 100 ns of molecular dynamics simulations at 30 °C. Larger pockets are colored in increasing intense reds. **[Con0, Con100]** Number of contacts made by residues in alpha fold model (0 ns), and following 100 ns of molecular dynamics simulations at 30 °C.

Res #	dSSP	AA	Thh Est Mut	ddG	INSEQs	GAPSEQs	PROSS	ASR	Lit. Stab	Lit. Act	Camsol SASA	RMSF 30	RMSF 45	RMSF 60	RMSF 75	NL	wD	nD	Camsol	AGGRESCAN	Pocket 0ns Vol	Pocket 100 ns	Con 0	Con 100
1	-	N	N		N	N					83.67	1.48	2.11	4.50	7.51	2.18	1.35	5.25	0.65	-0.83	0	0	3	3
2	-	P	P		P	P					99.40	1.15	4.77	3.98	6.41	1.64	1.10	3.95	1.43	-0.83	0	0	3	5
3	-	Y	Y		Y	Y					71.55	0.90	3.94	3.89	5.20	1.35	0.89	2.81	0.62	-0.83	0	65	5	5
4	-	E	E		Q	E					67.77	0.84	3.31	3.72	3.90	1.58	1.09	1.89	0.87	-0.83	0	65	4	5
5	-	R	R		R	R					18.66	0.87	2.65	3.37	3.01	2.21	1.45	1.25	0.06	-0.69	6	65	3	5
6	-	G	G		G	G					17.26	1.16	2.24	3.18	2.92	2.50	2.08	1.45	0.39	-0.56	6	0	3	3
7	-	P	P		P	P					81.44	1.39	1.97	2.75	2.49	2.20	1.60	1.69	1.58	-0.68	6	65	3	4
8	-	D	N	0.19	D	N					104.17	1.16	1.45	2.17	1.86	1.83	1.32	1.52	2.14	-0.84	0	0	3	3
9	-	P	P		P	P					11.91	0.98	1.12	1.78	1.39	1.41	1.07	0.99	0.06	-0.72	6	65	4	4

10	-	T	T		T	T					69.39	0.99	1.11	1.59	1.37	1.04	0.94	0.87	1.05	-0.38	0	0	5	5
11	-	E	N	-1.32	V	D					71.87	1.03	1.02	1.45	1.47	0.94	0.89	0.79	0.62	-0.47	0	4	5	5
12	-	S	S		S	A					78.92	0.96	0.95	1.29	1.52	0.88	0.88	0.77	1.11	-0.44	0	0	7	6
13	-	S	S		S	L					35.19	0.90	0.89	1.11	1.24	0.86	0.84	0.81	0.39	-0.08	6	65	8	7
14	-	I	I		L	L					0.57	0.77	0.71	1.03	0.97	0.75	0.67	0.63	0.00	-0.18	6	65	7	7
15	-	E	E		E	E					68.36	0.76	0.70	1.13	1.07	0.78	0.67	0.55	0.31	-0.23	33	0	7	6
16	-	A	A		A	A					33.96	0.80	0.77	1.08	1.10	0.84	0.75	0.68	0.24	-0.11	0	0	7	6
17	-	L	L		S	R					129.89	0.71	0.72	1.02	1.09	0.79	0.71	0.67	0.61	0.12	0	19	4	4
18	-	R	R		R	S					117.92	0.71	0.76	0.98	1.14	0.78	0.73	0.68	0.38	0.12	0	19	6	6
19	-	G	G		G	G					17.97	0.63	0.69	0.84	1.01	0.71	0.68	0.68	0.04	0.10	6	0	6	6
20	-	P	P		P	P					79.31	0.65	0.71	0.83	0.97	0.63	0.71	0.75	0.22	0.05	6	65	3	4
21	-	F	F		Y	F					21.39	0.56	0.63	0.73	0.86	0.56	0.63	0.68	0.04	-0.10	6	65	6	7
22	-	S	R	0.72	S	S					82.14	0.61	0.66	0.80	0.92	0.62	0.68	0.79	0.71	-0.41	0	0	5	6
23	B	V	V		V	V					35.77	0.55	0.60	0.74	0.81	0.58	0.60	0.68	0.29	-0.31	0	0	7	7
24	B	D	D		A	S					68.78	0.54	0.61	0.71	0.79	0.56	0.57	0.62	0.68	-0.07	0	0	5	5
25	B	E	E		T	E					101.98	0.61	0.66	0.77	0.88	0.60	0.59	0.72	0.95	-0.07	0	0	5	6
26	B	E	E		F	E					50.05	0.56	0.59	0.65	0.77	0.56	0.51	0.62	0.36	-0.29	0	145	5	5
27	B	S	R	0.80	N	N					94.37	0.66	0.66	0.72	0.84	0.63	0.60	0.73	1.60	-0.11	0	145	4	4
28	B	V	V		V	V				1.00	11.22	0.68	0.68	0.74	0.85	0.59	0.64	0.80	0.01	-0.29	0	145	6	6
29	-	S	S		S	S					59.90	1.05	1.00	1.09	1.34	0.84	0.95	1.43	-0.01	0.09	0	145	6	4
30	-	S	R	0.75	S	R					63.50	1.19	1.17	1.36	1.81	1.01	1.07	1.44	-0.55	0.22	0	0	7	6
31	-	L	L		L	L					173.20	1.73	1.60	1.91	2.36	1.39	1.46	1.69	-0.65	0.31	0	145	5	6
32	-	A	Q	0.49	S	S		S, H, T			62.10	2.43	1.65	1.92	3.07	1.47	1.49	1.99	-0.08	0.54	0	0	5	4
33	-	V	A	6.95	V	A					18.35	2.14	1.39	1.51	3.08	1.26	1.27	1.82	-0.04	0.30	7	145	5	4
34	-	S	R	0.17	S	S					82.19	2.23	1.65	1.64	2.62	1.76	1.48	1.68	-0.70	0.28	7	145	4	6
35	-	G	G		G	G					39.33	1.97	1.36	1.24	2.01	1.49	1.45	1.94	-0.01	0.25	0	8	5	4
36	B	F	F		F	F					5.20	1.53	1.22	1.15	1.59	1.21	1.21	1.32	0.00	0.08	7	8	8	7
37	B	G	G		G	G					13.45	1.14	1.16	1.05	1.18	0.99	1.00	1.10	0.02	0.29	0	145	7	8
38	B	G	G		G	G					0.36	0.72	0.68	0.67	0.84	0.67	0.73	0.77	0.00	0.24	0	0	9	8
39	B	G	G		G	G					8.24	0.54	0.54	0.59	0.68	0.51	0.57	0.54	-0.05	0.40	0	0	7	7
40	B	T	T		T	T					12.28	0.43	0.45	0.48	0.55	0.40	0.43	0.43	-0.02	0.42	0	0	8	8

41	B	I	I		I	I				1.28	0.41	0.45	0.47	0.53	0.39	0.39	0.43	0.00	0.21	0	145	7	7
42	B	Y	Y		Y	Y				18.36	0.40	0.46	0.50	0.55	0.40	0.40	0.44	-0.03	0.07	0	0	10	8
43	B	Y	Y		Y	Y				23.24	0.48	0.55	0.64	0.68	0.46	0.52	0.60	0.00	0.11	0	0	7	8
44	B	P	P		P	P				0.00	0.57	0.61	0.71	0.74	0.53	0.56	0.67	0.00	0.02	0	0	7	9
45	-	T	T		T	R				64.87	0.96	0.95	0.96	1.01	0.66	0.89	0.98	0.38	-0.12	0	0	5	5
46	-	D	D		N	E				61.98	1.24	1.38	1.54	1.45	0.83	1.12	1.24	0.63	-0.38	0	0	5	3
47	-	T	N	2.10	A	N				59.17	1.52	1.60	1.94	1.68	1.19	1.74	1.18	0.82	-0.53	0	0	6	5
48	-	N	N		G	S	S			134.97	1.47	1.71	2.16	2.11	1.30	1.44	1.77	1.08	-0.46	0	0	5	4
49	-	E	-		T	Q				86.77	1.44	1.69	1.88	1.92	1.31	1.23	1.49	0.63	-0.48	0	0	4	4
50	-	G	-		G	N				25.29	1.26	1.14	1.45	1.47	1.51	0.96	1.18	0.19	-0.47	0	0	5	5
51	-	T	T		T	T				51.05	0.62	0.68	0.86	0.95	0.83	0.61	0.88	0.39	-0.09	0	0	7	6
52	-	F	F		M	F				0.01	0.43	0.50	0.56	0.60	0.47	0.44	0.70	0.00	-0.07	0	0	7	7
53	B	G	G		A	G				0.00	0.40	0.49	0.53	0.57	0.44	0.42	0.59	0.00	0.27	0	0	8	7
54	B	A	A		A	A				0.00	0.41	0.48	0.53	0.54	0.43	0.43	0.47	0.00	0.43	0	0	7	6
55	B	V	V		I	V				0.00	0.38	0.43	0.47	0.48	0.36	0.38	0.41	0.00	0.45	0	0	7	7
56	B	A	A		A	A				0.00	0.35	0.40	0.42	0.44	0.33	0.36	0.37	0.00	0.41	0	0	8	6
57	B	I	I		V	I				0.30	0.36	0.39	0.43	0.45	0.34	0.38	0.38	0.00	0.34	0	0	7	8
58	B	A	S	0.03	V	S				0.00	0.40	0.44	0.47	0.49	0.36	0.45	0.40	0.00	0.38	80	43	6	6
59	-	P	P		P	P				0.09	0.51	0.52	0.57	0.60	0.40	0.54	0.57	0.00	0.38	80	43	9	10
60	-	G	G		G	G				2.13	0.85	0.80	0.70	0.94	0.58	0.65	1.12	0.00	0.17	80	43	8	7
61	-	Y	Y		F	Y				88.77	1.04	1.01	0.96	1.29	1.30	0.84	1.04	0.05	0.04	80	43	7	4
62	-	T	T		L	T			-1.00	77.97	1.06	1.02	1.31	1.75	1.41	0.94	0.94	0.10	-0.20	0	0	5	3
63	-	A	G	-0.71	A	G			1.00	26.65	0.74	0.91	1.31	2.00	1.45	0.80	0.74	-0.06	-0.22	80	0	6	4
64	H	S	T	0.09	A	T				27.69	0.65	0.79	0.99	2.31	0.69	0.83	0.60	-0.08	-0.09	1	0	7	6
65	H	Q	Q		E	Q				29.87	0.57	0.71	0.77	2.22	0.74	0.70	0.58	-0.02	-0.06	0	0	9	9
66	H	S	S		S	A				61.92	0.68	0.75	0.87	1.72	0.81	0.86	0.83	0.08	-0.07	0	0	6	5
67	H	S	S		S	S				23.94	0.68	0.68	0.95	1.38	0.78	0.93	0.83	-0.03	0.10	0	43	8	6
68	H	M	I	0.06	I	I				0.14	0.56	0.53	0.78	0.97	0.63	0.76	0.69	0.00	0.04	80	43	8	9
69	H	S	S		E	A				31.30	0.58	0.53	0.73	0.95	0.62	0.72	0.71	-0.11	-0.08	0	0	7	8
70	H	W	W		W	W				28.65	0.58	0.49	0.68	0.95	0.59	0.66	0.66	0.02	-0.08	0	19	6	5
71	H	L	L		W	L			-1.00	0.00	0.54	0.47	0.63	0.76	0.53	0.55	0.55	0.00	0.10	0	0	8	6

72	H	G	G		G	G				0.00	0.57	0.52	0.69	0.82	0.52	0.57	0.67	0.00	0.13	0	0	9	9
73	H	E	E		P	E				59.36	0.55	0.53	0.65	0.82	0.51	0.56	0.63	0.28	0.00	0	19	8	6
74	H	R	R		R	R				4.61	0.49	0.50	0.58	0.72	0.54	0.52	0.55	0.00	-0.11	0	19	9	9
75	H	L	L		L	I				0.00	0.42	0.45	0.52	0.62	0.50	0.45	0.46	0.00	-0.28	0	0	8	7
76	H	A	A		A	A				0.00	0.38	0.45	0.50	0.59	0.44	0.42	0.40	0.00	-0.24	0	0	12	11
77	H	S	S		S	S				0.93	0.41	0.48	0.54	0.65	0.43	0.47	0.46	0.00	0.00	0	0	11	13
78	H	Q	H	2.32	H	H				5.09	0.43	0.51	0.57	0.66	0.45	0.52	0.48	0.00	0.33	6	65	6	6
79	-	G	G		G	G				0.22	0.44	0.52	0.55	0.63	0.45	0.49	0.44	0.00	0.66	0	0	5	7
80	-	F	F		F	F				0.00	0.36	0.44	0.48	0.52	0.38	0.39	0.37	0.00	0.49	0	0	8	8
81	H	V	V		V	V				0.00	0.35	0.42	0.45	0.48	0.36	0.35	0.37	0.00	0.70	0	0	7	5
82	H	V	V		V	V				0.00	0.33	0.39	0.42	0.46	0.34	0.34	0.38	0.00	0.53	0	0	11	9
83	H	F	M	5.83	I	I				0.00	0.35	0.39	0.42	0.47	0.35	0.36	0.40	0.00	0.65	0	0	7	9
84	H	T	T		T	T				0.05	0.39	0.41	0.45	0.50	0.38	0.40	0.41	0.00	0.56	0	0	8	7
85	H	I	I		I	I				0.12	0.43	0.43	0.47	0.53	0.41	0.46	0.42	0.00	0.35	0	0	8	9
86	H	D	D		D	D				40.21	0.55	0.56	0.58	0.69	0.57	0.60	0.59	-0.03	0.04	1	0	8	7
87	-	T	T		T	T				10.41	0.90	0.78	0.94	0.95	0.93	0.73	1.00	0.01	0.01	1	0	7	6
88	-	N	N		N	N				85.87	1.50	1.23	1.46	1.60	1.79	1.10	2.13	0.12	-0.39	0	0	5	7
89	-	T	T		S	T				71.25	1.20	0.98	1.33	1.61	1.29	0.89	1.14	0.66	-0.51	0	0	6	5
90	-	R	T	-2.02	G	T				139.53	1.62	1.14	1.44	1.64	1.47	0.99	1.21	0.41	-0.75	1	0	7	5
91	-	L	L		F	L	F			106.23	1.51	1.14	1.30	1.33	1.56	0.87	1.28	0.55	-0.75	0	0	6	3
92	-	D	D		D	D				23.34	0.99	0.86	1.00	1.01	1.41	0.72	1.02	0.20	-0.76	0	0	5	5
93	-	Q	Q		Q	Q				91.84	0.63	0.63	0.73	0.75	1.11	0.59	0.61	0.97	-0.75	33	0	5	5
94	-	P	P		P	P				5.53	0.51	0.53	0.59	0.63	0.93	0.51	0.49	0.00	-0.80	33	27	8	9
95	H	D	D		A	D		-1.00	1.00	85.05	0.59	0.63	0.70	0.73	0.91	0.59	0.52	1.80	-0.86	33	27	8	7
96	H	S	S		S	S				44.20	0.57	0.63	0.68	0.73	0.77	0.58	0.55	1.05	-1.15	0	0	7	7
97	H	R	R		R	R				0.12	0.47	0.53	0.59	0.61	0.60	0.50	0.53	0.00	-0.80	0	0	8	9
98	H	G	A	1.59	A	A				1.62	0.50	0.53	0.58	0.59	0.58	0.53	0.52	0.00	-0.69	0	3	11	10
99	H	D	S	0.55	S	R				88.67	0.57	0.58	0.62	0.66	0.60	0.59	0.56	1.43	-0.66	0	3	7	7
100	H	Q	Q		Q	Q				3.12	0.52	0.57	0.61	0.63	0.52	0.56	0.58	0.00	-0.46	0	0	9	8
101	H	L	L		L	L		-1.00		0.00	0.50	0.55	0.61	0.60	0.50	0.55	0.59	0.00	-0.27	0	0	8	7
102	H	S	D	-0.14	M	N	L			22.97	0.60	0.61	0.63	0.67	0.57	0.65	0.57	0.16	-0.34	0	3	7	7

103	H	A	A		A	A				24.02	0.54	0.57	0.57	0.63	0.51	0.58	0.53	0.10	-0.15	7	8	8	8
104	H	A	A		A	A				0.00	0.49	0.55	0.58	0.62	0.45	0.53	0.55	0.00	0.21	0	0	7	7
105	H	L	L		L	L				0.00	0.60	0.62	0.65	0.71	0.51	0.66	0.60	0.00	0.52	0	5	7	7
106	H	D	D		D	D			-1.00	49.67	0.55	0.59	0.62	0.69	0.49	0.64	0.58	-0.24	0.21	0	0	8	9
107	H	Y	Y		Y	Y				17.54	0.55	0.60	0.63	0.71	0.47	0.60	0.54	0.03	0.04	7	145	8	9
108	H	L	M	0.11	L	M			-1.00	0.07	0.74	0.73	0.85	0.90	0.54	0.79	0.63	0.00	-0.10	0	145	7	5
109	H	V	V		V	I				29.73	0.77	0.81	0.92	0.98	0.61	0.87	0.73	0.23	-0.16	0	0	7	6
110	-	E	E		S	N			1.00	76.91	1.00	1.00	1.25	1.30	0.67	0.87	1.00	0.32	-0.45	3	0	6	6
111	-	D	D		Q	-				117.33	1.72	1.53	1.93	2.28	0.65	1.32	1.48	1.39	-0.28	3	8	4	5
112	-	N	-		N	-				63.58	1.74	1.75	2.42	2.63	-	1.69	1.96	0.86	-0.44	7	145	5	7
113	-	G	-		G	-				77.57	1.87	2.23	2.52	2.89	-	1.76	1.95	2.18	-0.62	3	145	5	4
114	-	R	-		T	R				135.37	1.43	1.72	1.90	2.24	-	1.56	1.85	1.25	-0.81	7	145	4	4
115	-	S	S		S	A				12.43	1.17	1.86	1.52	1.94	0.67	1.08	1.33	0.10	-0.47	3	145	6	6
116	-	S	S		S	S				37.16	0.94	1.13	1.07	1.26	0.84	0.76	0.97	0.23	-0.41	0	145	5	5
117	-	S	Y	0.22	S	S				67.37	0.91	0.98	1.01	1.18	0.82	0.73	1.07	1.14	-0.41	3	0	5	6
118	-	A	S	-0.33	P	A				48.91	0.84	0.98	0.93	1.13	0.86	0.75	1.02	0.38	-0.48	0	0	6	6
119	-	V	V		I	V				0.57	0.62	0.74	0.75	0.88	0.71	0.59	0.73	0.00	-0.14	0	145	7	7
120	-	R	R		Y	R				59.90	0.66	0.76	0.79	0.90	0.63	0.60	0.75	0.23	-0.31	3	145	9	7
121	-	N	N		G	S				95.25	0.73	0.82	0.86	0.97	0.65	0.67	0.90	0.99	-0.32	0	0	6	6
122	-	R	R		K	R				44.72	0.60	0.67	0.70	0.79	0.61	0.57	0.81	0.27	-0.49	0	0	7	7
123	-	I	I		V	I				9.02	0.55	0.61	0.65	0.72	0.58	0.53	0.64	0.00	-0.62	0	145	9	7
124	-	D	D		D	D				20.25	0.52	0.60	0.64	0.75	0.55	0.52	0.66	0.26	-0.65	0	0	8	8
125	-	P	S	-1.24	T	S				75.39	0.59	0.65	0.70	0.81	0.59	0.59	0.65	1.16	-0.51	4	5	6	4
126	-	D	S	1.40	N	S				101.45	0.56	0.65	0.68	0.74	0.59	0.57	0.70	1.28	-0.19	4	5	4	5
127	-	R	R		R	R				4.02	0.40	0.50	0.51	0.55	0.41	0.40	0.52	0.00	0.05	0	0	8	6
128	B	L	L		L	L			-1.00	5.20	0.38	0.45	0.48	0.49	0.37	0.39	0.47	0.00	-0.22	4	5	8	7
129	B	A	A		G	A				0.00	0.34	0.40	0.43	0.43	0.34	0.35	0.40	0.00	-0.13	0	0	8	7
130	B	V	A	-1.71	V	V				0.05	0.34	0.38	0.42	0.42	0.34	0.35	0.37	0.00	-0.12	0	0	8	7
131	B	M	M		M	M				0.00	0.33	0.36	0.42	0.41	0.32	0.35	0.36	0.00	0.18	0	0	8	9
132	B	G	G		G	G				0.00	0.37	0.38	0.43	0.45	0.36	0.38	0.37	0.00	0.26	80	43	11	10
133	-	H	H		W	H			-1.00	6.22	0.38	0.36	0.41	0.47	0.39	0.38	0.36	0.00	0.05	0	0	10	8

134	-	S	S		S	S					7.51	0.39	0.37	0.41	0.46	0.40	0.38	0.38	0.00	-0.01	80	43	9	9
135	-	M	M		M	M					24.61	0.43	0.41	0.47	0.51	0.43	0.41	0.42	0.05	-0.24	80	0	9	9
136	H	G	G		G	G					0.00	0.40	0.40	0.45	0.50	0.42	0.40	0.42	0.00	-0.36	80	43	11	13
137	H	G	G		G	G					0.00	0.38	0.38	0.42	0.46	0.41	0.38	0.40	0.00	-0.15	0	0	12	12
138	H	G	G		G	G					0.00	0.40	0.41	0.45	0.48	0.42	0.40	0.41	0.00	-0.04	33	27	11	11
139	H	G	G		G	G					0.00	0.40	0.41	0.45	0.48	0.40	0.41	0.41	0.00	-0.01	0	0	12	13
140	H	T	T		T	T			1.00		0.00	0.37	0.38	0.43	0.45	0.37	0.38	0.40	0.00	-0.11	0	0	12	11
141	H	L	L		L	L					0.00	0.39	0.41	0.46	0.47	0.40	0.40	0.41	0.00	-0.21	0	27	7	7
142	H	A	R	0.71	R	R		E, R			4.40	0.42	0.45	0.49	0.51	0.42	0.44	0.42	0.00	-0.36	33	27	9	8
143	H	A	L	3.11	A	L			-1.00		0.79	0.44	0.47	0.50	0.52	0.43	0.46	0.44	0.00	-0.43	0	3	10	10
144	H	A	A		A	A					0.20	0.47	0.50	0.54	0.55	0.51	0.50	0.48	0.00	-0.41	0	4	9	7
145	-	E	E		A	S	R/A				50.16	0.50	0.54	0.58	0.59	0.55	0.51	0.48	0.71	-0.43	0	4	7	7
146	-	D	R	0.72	D	Q	K				97.72	0.54	0.57	0.59	0.63	0.54	0.55	0.54	2.08	-0.43	0	3	5	5
147	-	R	R		G	R					50.57	0.49	0.56	0.55	0.59	0.49	0.55	0.52	0.38	-0.53	4	5	7	7
148	-	P	P		P	P					83.83	0.53	0.73	0.64	0.66	0.61	0.73	0.61	1.72	-0.53	0	4	7	4
149	-	S	D	0.17	R	D					72.34	0.54	0.67	0.64	0.67	0.57	0.67	0.60	1.47	-0.53	4	5	4	4
150	-	L	L		I	L					4.08	0.41	0.48	0.51	0.51	0.45	0.44	0.49	0.00	-0.17	4	5	7	7
151	B	K	Q	-0.39	K	K					79.68	0.37	0.43	0.48	0.47	0.41	0.38	0.43	0.39	0.00	0	0	8	7
152	B	A	A		A	A					0.03	0.36	0.41	0.45	0.44	0.39	0.36	0.39	0.00	0.29	0	0	8	8
153	B	A	A		A	A					0.02	0.33	0.37	0.42	0.40	0.36	0.34	0.35	0.00	0.31	0	0	7	6
154	B	I	I		I	I			1.00		0.01	0.32	0.34	0.39	0.38	0.33	0.34	0.32	0.00	0.31	0	0	7	5
155	B	P	P		P	P					0.00	0.34	0.34	0.40	0.40	0.37	0.37	0.36	0.00	0.27	0	0	9	9
156	-	L	L		L	L					0.00	0.36	0.34	0.41	0.42	0.48	0.40	0.40	0.00	0.26	80	43	8	8
157	-	T	T		A	T	A				0.77	0.42	0.37	0.44	0.46	0.53	0.45	0.52	0.00	0.41	80	43	7	6
158	-	P	P		P	P					0.12	0.41	0.37	0.42	0.43	0.50	0.45	0.53	0.00	0.21	13	13	9	8
159	-	W	W		W	W					47.41	0.43	0.41	0.45	0.47	0.52	0.44	0.46	-0.15	-0.09	80	27	7	7
160	-	H	H		D	H			1.00		15.54	0.48	0.50	0.52	0.55	0.64	0.51	0.54	0.00	-0.07	0	0	5	6
161	-	L	T	-0.72	S	L					144.11	0.60	0.68	0.70	0.76	0.90	0.71	0.64	-0.21	-0.11	33	27	3	3
162	-	D	D		S	N					46.19	0.61	0.74	0.78	0.82	0.96	0.77	0.57	-0.24	-0.13	33	27	5	5
163	-	K	K		S	K					58.20	0.62	0.70	0.78	0.79	0.95	0.75	0.64	0.00	-0.25	13	13	4	4
164	-	T	T		N	N					93.73	0.65	0.77	0.88	0.88	0.99	0.73	0.67	0.60	-0.18	0	0	6	6

165	-	W	W		F	W				1.68	0.59	0.70	0.79	0.79	0.98	0.62	0.61	0.00	-0.21	33	27	7	7
166	-	S	G	-0.46	S	S				70.92	0.71	0.80	0.87	0.94	0.96	0.75	0.64	0.39	-0.18	0	0	8	8
167	-	E	S	0.24	S	S		1.00		112.30	0.77	0.81	0.88	0.88	0.86	0.80	0.69	1.16	-0.02	0	4	4	4
168	-	V	V		I	V				0.00	0.61	0.68	0.73	0.70	0.73	0.63	0.62	0.00	0.07	0	4	7	7
169	-	R	R		T	R				140.30	0.52	0.61	0.69	0.62	0.65	0.54	0.54	0.70	0.24	0	4	4	3
170	-	V	V		V	V				0.18	0.39	0.47	0.52	0.48	0.53	0.41	0.41	0.00	0.33	0	4	8	6
171	-	P	P		P	P				23.74	0.35	0.42	0.49	0.45	0.39	0.36	0.37	0.04	0.56	0	0	6	6
172	B	T	T		T	T				0.00	0.34	0.38	0.45	0.41	0.38	0.36	0.34	0.00	0.66	0	0	8	6
173	B	L	L		L	L				0.15	0.33	0.36	0.42	0.40	0.37	0.35	0.33	0.00	0.48	0	0	7	7
174	B	I	I		I	I				0.00	0.34	0.35	0.41	0.41	0.41	0.37	0.34	0.00	0.46	0	0	6	5
175	B	I	I		F	I		1.00		0.20	0.38	0.39	0.44	0.45	0.49	0.41	0.37	0.00	0.14	0	0	8	7
176	B	G	G		A	G				0.15	0.44	0.42	0.47	0.49	0.54	0.46	0.39	0.00	-0.03	0	0	6	6
177	B	A	A		C	A			-1.00	0.49	0.52	0.44	0.56	0.59	0.74	0.48	0.42	0.00	-0.03	0	0	9	10
178	-	E	E		E	E				63.87	0.65	0.61	0.67	0.81	0.86	0.59	0.48	0.55	0.02	0	0	8	9
179	-	N	N		N	L				86.54	0.75	0.70	0.77	0.85	0.96	0.66	0.55	0.31	-0.19	80	43	8	7
180	-	D	D		D	D				6.86	0.68	0.65	0.78	0.87	0.81	0.75	0.51	0.00	-0.42	80	43	10	9
181	-	T	T		S	T			-1.00	131.54	0.81	0.84	0.90	1.10	0.91	0.99	0.61	0.23	-0.18	80	43	6	5
182	-	I	I		I	I				90.01	0.82	0.72	0.75	0.85	0.83	0.80	0.59	-0.26	-0.32	80	43	5	5
183	-	A	A		A	A				1.48	0.86	0.57	0.65	0.75	0.81	0.63	0.63	0.00	-0.19	80	43	6	4
184	-	S	S		P	P	P			40.45	0.57	0.53	0.60	0.62	0.87	0.57	0.49	-0.02	-0.16	13	13	9	8
185	-	V	V		V	V				18.84	0.53	0.52	0.55	0.58	0.83	0.55	0.53	0.02	0.01	0	13	7	7
186	H	R	R		S	A				203.59	0.56	0.62	0.62	0.63	0.77	0.59	0.77	0.28	-0.13	0	0	6	6
187	H	S	S		S	T				53.81	0.53	0.62	0.62	0.63	0.68	0.60	0.81	0.45	-0.37	13	13	5	5
188	H	H	H		H	H		1.00		16.21	0.44	0.48	0.49	0.51	0.51	0.50	0.55	0.03	-0.17	0	0	9	9
189	H	S	S		A	A				0.89	0.43	0.46	0.48	0.51	0.49	0.50	0.46	0.00	-0.01	13	13	7	6
190	H	E	E		S	K				38.04	0.45	0.56	0.55	0.60	0.52	0.58	0.62	0.11	-0.33	0	1	8	8
191	H	P	P		P	P				36.06	0.47	0.58	0.60	0.63	0.54	0.61	0.68	0.07	-0.23	0	0	7	7
192	H	F	F		F	F				0.00	0.43	0.49	0.53	0.54	0.49	0.51	0.47	0.00	-0.04	33	27	9	9
193	H	Y	Y		Y	Y				17.25	0.45	0.51	0.55	0.57	0.50	0.52	0.48	0.02	0.04	0	1	7	7
194	H	N	N		N	N	D, E			93.02	0.51	0.61	0.66	0.68	0.61	0.61	0.57	0.26	0.01	0	0	6	6
195	H	S	S		S	S			-1.00	28.47	0.52	0.63	0.70	0.69	0.70	0.60	0.55	0.12	0.13	0	0	8	8

196	H	L	L		I	L				5.05	0.54	0.63	0.72	0.69	0.71	0.58	0.56	0.00	0.32	0	0	7	7	
197	-	P	P		P	P				29.68	0.71	0.79	0.95	0.84	0.85	0.70	0.75	0.06	-0.08	0	0	6	6	
198	-	G	G		S	S	S			57.90	0.97	1.00	1.52	1.05	1.02	0.93	0.93	0.41	-0.31	0	0	4	4	
199	-	S	S		T	S				117.68	0.93	1.04	1.81	1.06	1.03	0.91	1.01	0.74	-0.17	0	0	4	5	
200	-	L	L		T	I				17.55	0.62	0.74	1.70	0.77	0.74	0.64	0.75	0.06	-0.01	0	0	6	6	
201	-	D	D		A	S				48.24	0.50	0.57	0.82	0.63	0.59	0.52	0.50	0.13	-0.01	0	40	6	6	
202	B	K	K		K	K				19.92	0.53	0.52	0.68	0.64	0.64	0.55	0.46	0.03	-0.13	0	40	9	8	
203	B	A	A		A	A				0.00	0.43	0.43	0.55	0.51	0.51	0.43	0.39	0.00	0.09	0	0	6	6	
204	B	Y	Y		F	Y		1.00		0.00	0.42	0.43	0.54	0.51	0.51	0.42	0.39	0.00	-0.08	0	1	8	8	
205	B	L	L		F	L				0.16	0.40	0.41	0.48	0.50	0.47	0.41	0.38	0.00	-0.30	0	0	6	6	
206	B	E	E		E	E				9.36	0.45	0.43	0.50	0.53	0.51	0.47	0.38	0.00	-0.10	0	0	9	9	
207	B	L	L		I	L				0.00	0.50	0.47	0.56	0.60	0.62	0.56	0.41	0.00	-0.03	0	0	9	9	
208	-	D	D		N	D	N		1.00	45.16	0.59	0.56	0.67	0.78	0.77	0.71	0.49	0.53	-0.14	0	0	8	8	
209	-	G	G		G	G			-1.00	39.25	0.67	0.66	0.85	0.98	1.03	0.81	0.55	0.82	-0.07	0	0	5	5	
210	-	A	A		G	A				0.63	0.61	0.60	0.82	0.95	0.96	0.69	0.51	0.00	-0.23	0	0	9	8	
211	-	S	S		S	T				55.71	0.65	0.65	1.01	1.05	0.97	0.73	0.51	-0.10	-0.11	80	43	8	8	
212	-	H	H		H	H				8.04	0.60	0.62	1.31	1.37	0.69	0.98	0.48	0.00	-0.41	0	0	5	6	
213	-	F	F		S	F		1.00	-1.00	119.31	0.66	0.68	1.33	1.45	0.78	0.97	0.58	-0.17	-0.05	80	43	6	7	
214	-	A	A		C	A			-1.00	4.23	0.66	0.60	1.80	1.58	1.71	0.74	0.72	0.00	-0.02	0	0	8	7	
215	-	P	P		A	P			-1.00	0.00	0.63	0.57	1.78	1.90	1.96	0.81	0.70	0.00	-0.05	80	43	4	4	
216	-	N	N		N	N				22.86	0.71	0.65	1.57	1.77	1.71	0.87	0.67	-0.06	-0.05	80	43	7	7	
217	-	L	L		G	I				97.64	1.15	1.08	1.71	2.01	2.06	1.34	0.86	0.07	0.19	0	0	6	6	
218	-	S	S		G	P	G			99.32	1.11	1.02	1.60	1.87	0.95	1.61	1.07	-0.05	-0.15	0	19	3	3	
219	-	S	-		N	N		N	-1.00	-1.00	25.53	1.10	1.07	1.38	1.60		1.56	1.07	-0.03	-0.16	0	0	7	4
220	-	S	-		S	K		ALL	-1.00	-1.00	103.84	0.98	1.01	1.17	1.40		1.19	0.92	-0.25	-0.14	0	0	4	4
221	-	Y	-		N	-		ALL			33.81	0.59	0.64	0.77	0.84		0.65	0.55	0.01	0.21	0	0	6	5
222	H	N	N		G	-				21.62	0.52	0.55	0.73	0.79	0.68	0.57	0.51	-0.06	0.05	0	19	8	7	
223	H	T	T		G	-				55.64	0.47	0.49	0.65	0.72	0.60	0.51	0.45	-0.23	-0.02	33	0	8	7	
224	H	T	T		L	I				15.06	0.44	0.46	0.58	0.60	0.56	0.47	0.42	-0.03	0.14	33	0	7	7	
225	H	I	I		L	I				0.18	0.41	0.43	0.54	0.53	0.50	0.45	0.40	0.00	0.14	0	0	7	6	
226	H	A	A		G	G				0.00	0.41	0.43	0.53	0.55	0.47	0.45	0.40	0.00	0.21	0	0	9	9	

227	H	K	K		K	K					28.41	0.42	0.45	0.53	0.56	0.44	0.45	0.40	-0.13	0.33	33	4	10	10
228	H	Y	Y		Y	Y					5.81	0.39	0.41	0.50	0.51	0.42	0.42	0.39	0.00	0.46	33	4	7	6
229	H	S	S		G	S					0.00	0.38	0.40	0.49	0.48	0.40	0.40	0.38	0.00	0.63	0	0	7	8
230	H	I	I		V	I					0.00	0.37	0.40	0.49	0.48	0.38	0.38	0.36	0.00	0.32	0	65	9	8
231	H	S	S		S	A					0.00	0.35	0.40	0.50	0.48	0.38	0.37	0.37	0.00	0.19	0	0	9	8
232	H	W	W		W	W					0.86	0.34	0.39	0.48	0.45	0.36	0.36	0.34	0.00	0.49	0	40	7	7
233	H	L	L		M	L					0.05	0.36	0.42	0.50	0.46	0.36	0.38	0.38	0.00	0.54	0	0	8	9
234	H	K	K		K	K					2.08	0.36	0.44	0.55	0.50	0.38	0.38	0.43	0.00	0.37	6	65	10	8
235	H	R	R		R	R					39.27	0.37	0.43	0.55	0.50	0.39	0.37	0.39	-0.02	-0.04	0	40	11	9
236	H	F	F		F	F					24.58	0.44	0.47	0.55	0.57	0.44	0.42	0.42	0.06	-0.21	0	40	8	7
237	H	V	V		L	V					2.45	0.45	0.47	0.55	0.59	0.43	0.42	0.46	0.00	-0.35	0	0	8	8
238	-	D	D		D	D					0.27	0.48	0.52	0.68	0.73	0.46	0.44	0.53	0.00	-0.64	0	65	8	4
239	-	D	D		N	N					70.54	0.52	0.56	0.98	0.79	0.51	0.47	0.54	0.85	-0.40	0	0	6	6
240	-	D	D		D	D					0.00	0.55	0.63	1.07	0.78	0.57	0.48	0.62	0.00	-0.30	0	65	8	6
241	H	T	T		T	T					85.23	0.69	0.83	1.45	0.90	0.80	0.64	0.68	1.17	-0.63	0	0	6	6
242	H	R	R		R	R					54.52	0.72	0.86	1.51	0.92	0.89	0.75	0.73	0.35	-0.61	0	65	5	5
243	H	Y	Y		Y	Y					8.03	0.60	0.73	1.12	0.83	0.78	0.64	0.58	0.00	-0.26	6	65	8	8
244	H	S	T	0.55	S	T					22.90	0.70	0.91	1.11	0.96	0.95	0.76	0.57	-0.10	0.02	0	40	8	8
245	H	Q	Q		Q	Q			-1.00		103.24	0.79	1.12	1.02	1.18	1.19	0.84	0.64	-0.45	0.18	0	0	6	6
246	H	F	F		F	F					16.45	0.76	1.07	0.90	1.13	1.10	0.87	0.60	-0.02	0.14	0	4	6	6
247	H	L	L		L	L					0.29	0.68	0.99	0.87	0.87	0.89	0.85	0.55	0.00	0.24	0	40	7	7
248	-	C	C		C	C					44.25	0.82	1.19	1.06	0.99	1.07	0.97	0.68	-0.09	0.08	0	0	8	6
249	-	P	P		G	P					115.32	1.28	1.39	1.36	1.43	1.38	1.21	0.89	-0.13	0.10	0	0	4	4
250	-	G	G		P	G					25.95	1.55	1.65	1.61	1.55	1.59	1.37	1.07	0.02	0.17	0	4	5	6
251	-	P	P		N	P					27.45	1.84	1.66	1.84	1.92	1.59	1.33	0.86	0.06	0.13	33	0	3	4
252	-	S	S		H	R					89.07	1.80	1.94	2.17	2.82	2.32	2.15	1.11	0.40	-0.08	0	4	3	6
253	-	T	T		E	D					68.60	2.59	2.75	2.90	3.76	3.34	3.47	1.27	0.33	-0.18	0	0	7	8
254	-	G	G		A	G					27.36	3.07	2.57	3.64	4.28	3.74	4.01	2.19	0.14	-0.35	0	0	5	8
255	-	L	W	0.08	D	L					194.35	2.98	1.97	3.42	4.37	3.67	3.70	2.32	0.50	-0.11	0	0	4	5
256	-	G	G		L	F					94.43	2.46	1.95	2.76	3.71	2.91	2.87	1.78	0.18	-0.23	33	4	3	5
257	-	S	S		T	G					18.46	1.61	1.49	1.69	2.21	2.29	1.80	1.06	0.12	-0.36	33	4	7	7

258	-	D	D		R	E				74.51	0.83	0.77	0.85	1.09	0.98	0.95	0.61	0.91	-0.21	33	0	5	6
259	B	V	V		I	V				3.55	0.64	0.64	0.68	0.82	0.72	0.76	0.52	0.00	-0.29	33	0	9	10
260	B	E	E		S	E				63.12	0.64	0.63	0.66	0.77	0.71	0.75	0.50	0.83	-0.47	0	0	8	8
261	B	E	E		E	E				43.50	0.60	0.64	0.64	0.71	0.67	0.70	0.47	0.44	-0.43	0	1	7	5
262	B	Y	Y		Y	Y				37.69	0.56	0.65	0.65	0.67	0.65	0.67	0.46	0.12	-0.33	0	0	5	5
263	B	R	R		R	R				73.90	0.54	0.65	0.72	0.71	0.68	0.62	0.49	0.22	-0.17	0	1	5	5
264	-	S	S		G	S				34.63	0.68	1.09	0.94	1.12	0.99	0.88	0.65	0.07	-0.46	0	40	6	6
265	-	T	T		N	T				52.57	1.00	1.31	1.53	1.50	1.63	1.37	0.99	-0.03	-0.46	0	40	5	4
266	-	C	C		C	C				57.99	1.19	1.41	1.61	1.70	1.50	1.59	0.85	0.00	-0.46	0	0	5	5
267	-	P	P		P	P				40.53	2.01	2.35	2.32	3.05	1.81	2.58	0.69	-0.11	-0.46	0	40	3	5
268	-	-	F		Y	F																	

N. ANC Cut-Therm Extended Residue Data

Table A4 Extended Residue Data for ANC Cut-Therm. The below table represents data collected for ANC Cut-Therm from several techniques at a residue level. **[Res #]** Refers to ANC residue #. **[dSSP]** Refers to secondary structure of ANC, where B (yellow) is beta sheet, and H (blue) is helix. **[AA]** Refers to ANC amino acid. **[Thh Est Mut]** The sequence for Thh_Est (closest ancestor) for comparison, where mutations are highlighted in red. **[ddG]** Predicted stability upon mutation of Thh Est to ancestor residue, where larger numbers are considered destabilizing (>1, red), and negative numbers are considered stabilizing (<-1, green). **[INSEQs, GAPSEQs]** Refers to known PET hydrolase INSEQ and GAPSEQ consensus residues. Changes from ancestor are colored in red. **[PROSS]** Mutations that are predicted to be stabilizing by PROSS (Goldenzeig, Goldsmith, ... Fleishman, 2016). **[ASR]** For sites with <50% likelihood for a given ancestral residue, alternative residues are listed in order of likelihood. **[Lit Stab, Lit Act]** Comparison of ancestor residues to known literature mutations, where a -1 refers to destabilizing or deactivating residue in the ancestor (red), and +1 refers to stabilizing or activating residue in ancestor (green). **[Camsol SASA]** Extent of residue burial predicted by Camsol, where low values suggest a buried residue (yellow) and high values suggest solvent exposure (blue) (Sormanni, Aprile, and Vendruscolo, 2015). **[RMSF 30, RMSF 45, RMSF 60, RMSF 75]** Residue fluctuation averages from molecular dynamics simulations, where higher fluctuations are coloured red. The number 30, 45, 60, and 75 refers to the temperature the simulation was run. **[NL, wD, nD]** Residue fluctuation averages from molecular dynamics simulations, where higher fluctuations are red. NL refers to no loop simulations, wD refers to simulations with the additional active site disulfide, and nD refers to simulations with a reduced terminal disulfide (no active site disulfide). **[Camsol, AGGRESCAN]** Aggregation propensity predictions by CAMSOL (structure corrected), and AGGRESCAN where larger values indicate higher aggregation propensity (red) (Conchillo-Solé, de Groot, ... Ventura, 2007; Sormanni, Aprile, and Vendruscolo, 2015). **[Pocket 0ns, Pocket 100 ns]** Residue involvement with pockets in alpha fold model (0 ns), compared to pocket involvement following 100 ns of molecular dynamics simulations at 30 °C. Larger pockets are colored in increasing intense reds. **[Con0, Con100]** Number of contacts made by residues in alpha fold model (0 ns), and following 100 ns of molecular dynamics simulations at 30 °C.

Res #	dSSP	AA	Thh_Est Mut	ddG	INSEQs	Gaps eq Consensus	PROSS	ASR	Lit. Stab	Lit. Act	SASA	RMSF 30	RMSF 45	RMSF 60	RMSF 75	NL	wD	nD	Camsol	AGGR ESCAN	Pocket 0ns Vol	Pocket 100 ns	Con 0	Con 100
1	-	N	N		N	N					79.10	1.14	5.03	3.86	6.71	2.80	2.85	2.27	0.59	-0.77	0	4	4	2
2	-	P	P		P	P					107.95	1.05	3.86	3.25	5.84	1.93	2.62	1.88	1.42	-0.77	0	4	4	3
3	-	Y	Y		Y	Y					69.89	0.87	3.09	3.33	5.04	1.11	1.90	1.50	0.62	-0.77	0	4	6	5
4	-	E	E		Q	E					73.51	1.01	3.03	3.05	4.09	0.69	1.25	1.78	0.94	-0.77	0	73	6	6
5	-	R	R		R	R					15.70	1.20	3.10	3.23	3.49	0.59	0.83	2.48	0.02	-0.63	45	73	4	5
6	-	G	G		G	G					21.42	1.27	2.70	3.30	2.96	0.73	0.81	2.58	0.41	-0.50	45	73	3	4
7	-	P	P		P	P					82.88	1.38	2.15	2.61	2.52	0.98	1.08	1.84	1.23	-0.62	45	73	3	3
8	-	N	N		D	N					108.66	1.26	1.54	2.05	2.03	1.05	1.14	1.41	1.56	-0.78	45	73	3	3

9	-	P	P		P	P					11.52	1.22	1.13	1.55	1.70	1.03	0.98	1.02	0.04	-0.66	45	73	4	3
10	-	T	T		T	T					68.94	1.47	1.33	1.58	1.77	1.12	1.04	1.12	0.78	-0.32	0	73	5	4
11	-	E	N	-1.32	V	D		D			73.27	1.58	1.30	1.69	1.93	1.19	1.23	1.19	0.43	-0.41	0	24	5	5
12	-	S	S		S	A					79.03	1.65	1.30	1.70	1.69	1.26	1.20	1.39	1.04	-0.38	0	0	6	7
13	-	S	S		S	L					36.50	1.53	1.26	1.45	1.49	1.15	1.07	1.32	0.38	-0.08	45	73	8	8
14	-	I	I		L	L					0.36	1.22	1.01	1.18	1.41	0.89	1.08	0.81	0.00	-0.18	45	73	7	6
15	-	E	E		E	E					88.01	1.16	1.12	1.32	1.71	1.07	1.10	0.73	0.41	-0.23	0	4	6	6
16	-	A	A		A	A					35.89	1.27	1.36	1.49	1.78	1.25	1.04	0.81	0.26	-0.11	45	0	6	4
17	-	L	L		S	R					137.4 2	1.01	1.16	1.39	1.82	1.24	0.97	0.85	0.58	0.12	1	0	3	5
18	-	R	R		R	S					167.4 2	0.89	1.04	1.32	1.41	1.25	0.88	0.70	0.50	0.12	45	73	6	5
19	-	G	G		G	G					17.83	0.71	0.85	1.08	1.03	1.09	0.71	0.57	0.04	0.10	45	73	6	7
20	-	P	P		P	P					80.41	0.71	0.82	0.91	0.92	0.97	0.69	0.66	0.28	0.05	45	73	3	4
21	-	F	F		Y	F					21.85	0.61	0.71	0.72	0.80	0.75	0.58	0.63	0.04	-0.10	45	0	6	6
22	-	S	R	0.72	S	S					80.36	0.67	0.77	0.75	0.86	0.68	0.62	0.74	0.67	-0.41	0	0	5	5
23	B	V	V		V	V					40.53	0.60	0.71	0.68	0.77	0.55	0.58	0.69	0.34	-0.41	2	0	7	5
24	B	D	D		A	S					67.46	0.56	0.69	0.65	0.75	0.50	0.57	0.69	0.88	-0.18	0	0	6	5
25	B	E	E		T	E					90.76	0.59	0.71	0.68	0.82	0.56	0.57	0.84	1.01	-0.17	20	0	5	5
26	B	E	E		F	E					47.09	0.52	0.65	0.61	0.74	0.51	0.51	0.67	0.43	-0.50	0	0	5	5
27	B	R	R		N	N					200.2 4	0.59	0.69	0.66	0.79	0.56	0.57	0.58	2.13	-0.32	20	0	4	4
28	B	V	V		V	V			1.00	9.48	0.61	0.69	0.72	0.79	0.59	0.58	0.54	0.01	-0.50	0	0	6	6	
29	-	S	S		S	S					61.21	0.87	0.96	1.11	1.19	0.88	0.84	0.65	0.57	-0.30	0	0	6	6
30	-	R	R		S	R					148.6 5	0.91	1.01	1.31	1.29	0.93	0.89	0.71	0.19	-0.18	0	0	7	7
31	-	L	L		L	L					175.5 6	1.00	1.15	1.57	1.73	1.02	0.97	0.85	0.11	-0.08	0	0	5	5
32	-	A	Q	0.49	S	S	Q	Q, G, S			61.98	0.95	1.14	1.59	2.33	0.96	0.91	0.82	0.32	0.26	0	0	5	5
33	-	A	A		V	A					15.03	0.78	1.01	1.45	2.39	0.78	0.74	0.71	0.03	0.02	0	0	6	6
34	-	S	R	0.17	S	S					89.70	1.24	1.44	2.09	2.43	1.14	1.17	1.14	-0.42	-0.01	0	0	4	4
35	-	G	G		G	G					43.68	1.08	1.12	1.88	2.01	0.97	0.96	0.82	0.12	0.07	0	0	5	5
36	B	F	F		F	F					3.58	0.70	0.81	1.43	1.53	0.68	0.66	0.60	0.00	-0.10	0	0	8	8
37	B	G	G		G	G					5.56	0.63	0.77	1.19	1.28	0.62	0.61	0.56	0.00	0.11	0	0	7	7
38	B	G	G		G	G					1.55	0.62	0.69	0.84	0.80	0.62	0.61	0.58	0.00	0.24	0	0	10	10
39	B	G	G		G	G					4.63	0.54	0.62	0.65	0.68	0.52	0.54	0.54	0.00	0.40	20	0	7	7

40	B	T	T		T	T					13.59	0.43	0.49	0.49	0.55	0.41	0.42	0.42	-0.02	0.42	20	0	8	8	
41	B	I	I		I	I					0.00	0.40	0.48	0.47	0.52	0.39	0.39	0.45	0.00	0.21	0	0	7	7	
42	B	Y	Y		Y	Y					18.64	0.41	0.50	0.49	0.53	0.39	0.40	0.49	-0.03	0.07	20	0	10	6	
43	B	Y	Y		Y	Y					9.47	0.48	0.59	0.57	0.64	0.45	0.50	0.59	0.00	-0.02	0	0	7	7	
44	B	P	P		P	P					0.12	0.57	0.67	0.63	0.75	0.54	0.56	0.65	0.00	-0.11	0	0	7	7	
45	-	T	T		T	R					60.99	0.89	0.99	0.94	1.22	0.83	0.73	0.93	0.39	-0.25	0	0	6	4	
46	-	D	D		N	E					63.77	0.97	1.08	1.13	1.42	1.16	0.85	1.07	0.86	-0.51	0	0	5	5	
47	-	N	N		A	N					55.45	1.31	1.60	1.74	2.02	1.36	1.28	1.92	0.86	-0.65	0	8	5	4	
48	-	N	N		G	S					130.75	1.38	1.51	1.80	2.14	1.53	1.34	1.94	1.39	-0.59	0	8	5	4	
49	-	E	-		T	Q					78.24	1.24	1.34	1.69	2.10	1.24	1.09	1.46	0.77	-0.61	0	8	4	4	
50	-	G	-		G	N					29.04	1.13	1.14	1.37	1.41	1.11	1.00	1.25	0.40	-0.60	0	8	5	3	
51	-	T	T		T	T					48.74	0.72	0.83	0.93	0.92	0.68	0.70	0.76	0.38	-0.22	0	8	6	6	
52	-	F	F		M	F					0.16	0.45	0.57	0.57	0.62	0.45	0.48	0.51	0.00	-0.07	0	0	7	6	
53	B	G	G		A	G					0.00	0.45	0.52	0.54	0.64	0.45	0.51	0.50	0.00	0.27	0	0	8	8	
54	B	A	A		A	A					0.01	0.43	0.49	0.52	0.62	0.45	0.50	0.49	0.00	0.40	0	0	7	7	
55	B	V	V		I	V					0.00	0.35	0.42	0.43	0.49	0.37	0.37	0.37	0.00	0.42	0	0	7	6	
56	B	A	A		A	A					0.00	0.34	0.39	0.40	0.44	0.34	0.34	0.35	0.00	0.38	0	0	7	6	
57	B	I	I		V	I					0.00	0.36	0.40	0.41	0.46	0.34	0.34	0.34	0.00	0.31	99	142	7	7	
58	B	S	S		V	S					0.00	0.39	0.43	0.44	0.54	0.34	0.37	0.34	0.00	0.35	99	142	6	6	
59	-	P	P		P	P					0.00	0.47	0.48	0.54	0.76	0.38	0.48	0.38	0.00	0.30	99	142	9	7	
60	-	G	G		G	G					4.65	0.56	0.62	0.87	0.93	0.48	0.66	0.49	0.00	0.10	99	142	8	8	
61	-	Y	Y		F	Y					106.21	0.72	1.02	1.42	1.16	0.61	0.83	1.03	-0.13	-0.03	99	142	7	5	
62	-	T	T		L	T					-1.00	81.75	0.98	1.48	1.97	1.65	0.79	0.93	1.43	-0.22	-0.27	0	142	5	5
63	-	G	G		A	G					-1.00	17.41	1.15	1.46	1.63	1.84	0.84	1.01	1.52	-0.10	-0.27	99	142	6	5
64	H	T	T		A	T					58.57	0.90	1.06	1.60	1.78	0.70	0.84	0.76	-0.24	-0.03	0	4	7	5	
65	H	Q	Q		E	Q					27.18	0.77	0.92	1.34	2.03	0.57	0.69	0.54	-0.04	0.00	20	0	10	8	
66	H	S	S		S	A					80.93	0.82	0.90	1.27	1.79	0.62	0.72	0.52	0.08	-0.01	20	0	6	6	
67	H	S	S		S	S					29.64	0.74	0.79	1.17	1.55	0.69	0.76	0.59	-0.17	0.16	99	142	7	4	
68	H	I	I		I	I					0.00	0.58	0.61	0.76	1.02	0.61	0.67	0.60	0.00	0.16	99	142	7	7	
69	H	S	S		E	A					37.47	0.61	0.65	0.71	0.95	0.56	0.66	0.61	-0.21	0.02	20	0	8	6	
70	H	W	W		W	W					49.88	0.69	0.71	0.73	0.93	0.53	0.70	0.65	0.02	0.02	1	142	5	6	

71	H	L	L		W	L				-1.00	0.00	0.62	0.63	0.65	0.78	0.51	0.63	0.61	0.00	0.21	0	142	7	8	
72	H	G	G		G	G					0.00	0.55	0.59	0.62	0.72	0.53	0.62	0.64	0.00	0.23	0	0	9	8	
73	H	E	E		P	E					69.51	0.58	0.61	0.60	0.76	0.52	0.65	0.67	0.29	0.00	2	0	8	6	
74	H	R	R		R	R					13.85	0.60	0.59	0.60	0.72	0.50	0.62	0.57	0.00	-0.08	45	73	8	7	
75	H	L	L		L	I					0.00	0.53	0.55	0.57	0.63	0.47	0.53	0.49	0.00	-0.26	45	73	8	7	
76	H	A	A		A	A					0.00	0.46	0.52	0.52	0.58	0.44	0.45	0.45	0.00	-0.22	0	0	12	9	
77	H	S	S		S	S					1.13	0.47	0.57	0.56	0.63	0.47	0.48	0.45	0.00	0.02	45	73	11	10	
78	H	H	H		H	H			1.00		4.42	0.51	0.58	0.59	0.67	0.52	0.51	0.45	0.00	0.35	0	0	6	5	
79	-	G	G		G	G					0.00	0.48	0.54	0.54	0.64	0.50	0.49	0.46	0.00	0.59	45	73	5	6	
80	-	F	F		F	F					0.00	0.39	0.47	0.46	0.55	0.40	0.40	0.41	0.00	0.42	45	73	8	6	
81	H	V	V		V	V					0.00	0.35	0.44	0.44	0.49	0.37	0.37	0.38	0.00	0.63	0	0	7	6	
82	H	V	V		V	V					0.00	0.34	0.43	0.43	0.45	0.37	0.36	0.38	0.00	0.46	0	0	11	9	
83	H	M	M		I	I			I, F		0.00	0.37	0.43	0.44	0.45	0.40	0.38	0.38	0.00	0.55	0	0	6	6	
84	H	T	T		T	T					0.00	0.39	0.45	0.46	0.49	0.39	0.38	0.38	0.00	0.47	0	0	8	6	
85	H	I	I		I	I					0.00	0.40	0.47	0.48	0.53	0.39	0.38	0.42	0.00	0.26	0	0	7	8	
86	H	D	D		D	D					35.44	0.47	0.52	0.69	0.71	0.45	0.45	0.46	-0.02	0.06	0	4	8	7	
87	-	T	T		T	T					6.19	0.51	0.69	1.09	1.18	0.49	0.52	0.50	0.00	0.04	0	4	7	7	
88	-	N	N		N	N					55.49	0.57	1.10	1.50	1.56	0.56	0.59	0.60	-0.17	-0.27	0	0	5	5	
89	-	T	T		S	T					69.39	0.71	1.11	1.26	1.22	0.65	0.70	0.88	0.01	-0.39	0	4	6	6	
90	-	T	T		G	T					51.53	0.98	1.36	1.72	1.22	0.82	0.94	1.36	-0.20	-0.63	0	4	7	5	
91	-	L	L		F	L					103.47	0.89	1.14	1.51	1.18	0.71	0.79	1.05	-0.14	-0.63	0	0	6	5	
92	-	D	D		D	D					26.56	0.64	0.84	1.06	0.98	0.53	0.57	0.70	0.03	-0.64	0	0	6	4	
93	-	Q	Q		Q	Q					90.47	0.56	0.65	0.78	0.72	0.49	0.50	0.61	0.27	-0.63	0	0	5	6	
94	-	P	P		P	P					5.38	0.47	0.57	0.65	0.61	0.43	0.45	0.56	0.00	-0.62	10	14	8	8	
95	H	D	D		A	D				-1.00	1.00	48.92	0.50	0.59	0.65	0.63	0.47	0.49	0.55	0.75	-0.64	10	0	7	7
96	H	S	S		S	S					45.66	0.46	0.57	0.61	0.61	0.44	0.46	0.50	0.62	-0.93	0	0	8	7	
97	H	R	R		R	R					0.00	0.40	0.52	0.54	0.54	0.38	0.40	0.44	0.00	-0.57	0	4	8	9	
98	H	A	A		A	A					0.00	0.39	0.47	0.51	0.52	0.38	0.40	0.40	0.00	-0.64	0	0	10	10	
99	H	S	S		S	R					58.27	0.42	0.47	0.54	0.56	0.41	0.44	0.42	0.70	-0.60	0	0	7	7	
100	H	Q	Q		Q	Q					2.96	0.40	0.47	0.52	0.55	0.40	0.42	0.40	0.00	-0.40	0	0	9	9	
101	H	L	L		L	L				-1.00	0.00	0.37	0.45	0.49	0.51	0.39	0.40	0.38	0.00	-0.22	0	0	7	7	

102	H	D	D		M	N					42.22	0.40	0.46	0.51	0.53	0.40	0.43	0.41	0.12	-0.28	0	0	7	7
103	H	A	A		A	A					22.18	0.41	0.46	0.51	0.53	0.41	0.43	0.42	0.10	-0.15	0	0	8	8
104	H	A	A		A	A					0.00	0.40	0.47	0.51	0.54	0.44	0.42	0.41	0.00	-0.02	0	0	7	7
105	H	L	L		L	L					0.04	0.41	0.52	0.56	0.57	0.48	0.47	0.44	0.00	0.30	0	0	7	7
106	H	D	D		D	D			-1.00		79.60	0.45	0.55	0.62	0.61	0.49	0.51	0.50	-0.29	-0.01	0	0	7	8
107	H	Y	Y		Y	Y					38.33	0.45	0.53	0.63	0.62	0.51	0.50	0.50	0.12	0.35	0	0	7	7
108	H	M	M		L	M					0.02	0.45	0.62	0.81	0.79	0.63	0.60	0.65	0.00	0.32	0	0	7	7
109	H	V	V		V	I					37.44	0.48	0.70	0.91	0.90	0.71	0.70	0.75	0.32	0.29	0	0	7	8
110	-	E	E		S	N			1.00		124.07	0.52	0.65	0.81	0.82	0.74	0.65	0.64	0.28	-0.07	0	0	7	8
111	-	L	-		Q	-					26.01	0.52	0.71	1.01	1.03		0.72	0.76	0.08	0.10	0	0	8	7
112	-	S	-		N	-		A, T	-1.00		7.69	0.54	0.84	1.30	1.36		0.89	1.09	0.01	-0.06	0	0	11	10
113	-	S	-		G	-		ALL	1.00		96.43	0.67	0.89	1.27	1.51		0.94	1.17	1.62	-0.19	0	0	6	6
114	-	D	D		T	R					62.18	0.82	1.03	1.43	1.73	0.85	1.04	1.36	0.59	-0.40	0	0	8	8
115	-	S	S		S	A					86.67	1.04	1.46	2.10	2.26	1.51	1.43	1.99	1.22	-0.07	0	0	5	6
116	-	S	S		S	S					97.74	1.05	1.40	1.95	2.22	1.49	1.39	1.75	0.55	-0.36	0	0	4	5
117	-	S	Y	0.22	S	S					13.22	0.82	1.05	1.65	2.02	1.32	1.17	1.47	0.08	-0.47	0	0	8	7
118	-	S	S		P	A					57.35	0.83	1.03	1.54	1.93	1.15	1.08	1.74	0.57	-0.58	0	0	4	4
119	-	V	V		I	V					0.02	0.70	0.86	0.98	1.39	0.89	0.80	1.38	0.00	-0.17	0	0	6	5
120	-	R	R		Y	R	Y				125.64	0.62	0.85	1.05	1.37	1.02	0.78	1.38	0.92	-0.34	0	0	10	9
121	-	N	N		G	S					99.82	0.62	0.89	1.05	1.36	0.90	0.78	1.22	1.42	-0.34	0	8	5	6
122	-	R	R		K	R					31.07	0.54	0.69	0.75	0.89	0.62	0.58	0.78	0.15	-0.34	0	8	8	4
123	-	I	I		V	I					16.02	0.52	0.67	0.67	0.79	0.57	0.58	0.66	0.09	-0.45	0	0	9	6
124	-	D	D		D	D					25.73	0.57	0.70	0.69	0.78	0.55	0.61	0.68	0.25	-0.47	0	0	8	8
125	-	S	S		T	S	P/A				66.33	0.66	0.76	0.77	0.83	0.63	0.67	0.69	0.80	-0.34	2	0	5	5
126	-	S	S		N	S					65.05	0.61	0.71	0.73	0.84	0.63	0.63	0.68	0.43	-0.02	2	0	4	4
127	-	R	R		R	R					4.96	0.41	0.52	0.53	0.62	0.45	0.43	0.49	0.00	0.22	0	0	8	7
128	B	L	L		L	L			-1.00		9.01	0.36	0.45	0.44	0.53	0.38	0.39	0.39	0.00	-0.04	2	0	8	7
129	B	A	A		G	A					0.01	0.34	0.40	0.39	0.47	0.34	0.35	0.35	0.00	0.05	0	0	8	9
130	B	V	A	-1.71	V	V					0.00	0.35	0.40	0.41	0.47	0.34	0.34	0.34	0.00	0.05	0	0	8	8
131	B	M	M		M	M					0.00	0.36	0.40	0.41	0.46	0.33	0.34	0.33	0.00	0.18	0	142	8	7
132	B	G	G		G	G					0.00	0.42	0.43	0.47	0.50	0.37	0.38	0.36	0.00	0.26	99	142	11	11

133	-	H	H		W	H				-1.00	7.17	0.43	0.44	0.50	0.52	0.36	0.40	0.35	0.00	0.05	0	0	10	9
134	-	S	S		S	S					7.93	0.41	0.45	0.49	0.49	0.36	0.42	0.39	0.00	-0.01	99	142	9	10
135	-	M	M		M	M					23.30	0.43	0.47	0.51	0.51	0.38	0.41	0.42	0.02	-0.24	99	0	9	11
136	H	G	G		G	G					0.00	0.44	0.45	0.48	0.50	0.38	0.40	0.42	0.00	-0.36	99	0	11	12
137	H	G	G		G	G					0.00	0.41	0.43	0.45	0.48	0.38	0.39	0.39	0.00	-0.15	0	0	12	12
138	H	G	G		G	G					0.00	0.43	0.45	0.48	0.50	0.40	0.42	0.42	0.00	-0.17	10	14	11	11
139	H	G	G		G	G					0.00	0.42	0.44	0.48	0.51	0.39	0.42	0.42	0.00	0.02	0	0	12	12
140	H	T	T		T	T			1.00		0.00	0.41	0.42	0.45	0.51	0.37	0.40	0.39	0.00	-0.09	0	0	11	11
141	H	L	L		L	L					0.00	0.43	0.45	0.48	0.56	0.40	0.43	0.40	0.00	-0.19	10	0	7	7
142	H	R	R		R	R					44.11	0.44	0.49	0.51	0.60	0.42	0.45	0.42	0.16	-0.27	10	14	8	7
143	H	L	L		A	L					2.42	0.43	0.49	0.52	0.60	0.44	0.47	0.42	0.00	-0.35	0	0	10	9
144	H	A	A		A	A					0.17	0.45	0.52	0.57	0.67	0.48	0.52	0.47	0.00	-0.32	0	0	8	9
145	-	E	E		A	S			E, S		47.37	0.50	0.60	0.61	0.76	0.53	0.58	0.50	0.77	-0.51	0	2	7	7
146	-	R	R		D	Q					114.7 2	0.55	0.65	0.67	0.80	0.58	0.64	0.54	1.33	-0.51	0	0	5	5
147	-	R	R		G	R					56.74	0.52	0.61	0.64	0.79	0.52	0.65	0.58	0.56	-0.47	0	0	8	7
148	-	P	P		P	P					54.44	0.54	0.66	0.76	1.20	0.61	0.80	0.63	1.32	-0.63	0	0	6	7
149	-	D	D		R	D					74.08	0.55	0.66	0.69	0.91	0.61	0.68	0.60	1.45	-0.63	2	0	4	4
150	-	L	L		I	L					3.32	0.41	0.48	0.48	0.62	0.42	0.44	0.42	0.00	-0.27	2	0	9	7
151	B	K	Q	-0.39	K	K					78.25	0.37	0.45	0.43	0.54	0.36	0.38	0.37	0.18	-0.17	0	0	8	6
152	B	A	A		A	A					0.00	0.36	0.42	0.41	0.50	0.36	0.37	0.36	0.00	0.12	0	0	8	7
153	B	A	A		A	A					0.00	0.34	0.39	0.38	0.47	0.34	0.34	0.34	0.00	0.14	0	0	7	7
154	B	I	I		I	I			1.00		0.03	0.33	0.37	0.37	0.43	0.32	0.33	0.31	0.00	0.31	0	0	7	6
155	B	P	P		P	P					0.00	0.38	0.39	0.41	0.46	0.34	0.37	0.33	0.00	0.27	0	0	9	11
156	-	L	L		L	L					0.00	0.46	0.44	0.47	0.53	0.37	0.42	0.33	0.00	0.26	99	142	8	8
157	-	T	T		A	T	A/Q				0.71	0.63	0.51	0.53	0.58	0.43	0.50	0.37	0.00	0.41	99	142	7	9
158	-	P	P		P	P					0.03	0.63	0.49	0.53	0.56	0.42	0.48	0.40	0.00	0.21	12	32	9	8
159	-	W	W		W	W					47.48	0.47	0.46	0.50	0.55	0.44	0.43	0.52	-0.15	-0.09	99	32	7	6
160	-	H	H		D	H			1.00		13.42	0.51	0.53	0.59	0.63	0.49	0.49	0.55	0.00	-0.07	0	0	5	4
161	-	L	T	-0.72	S	L					144.0 9	0.66	0.75	0.87	0.87	0.66	0.68	0.65	-0.36	-0.11	12	32	3	3
162	-	D	D		S	N					37.75	0.71	0.81	0.96	0.95	0.69	0.73	0.62	-0.24	-0.13	10	14	5	4
163	-	K	K		S	K					57.89	0.75	0.82	0.99	1.01	0.71	0.81	0.61	-0.01	-0.12	12	1	4	3

164	-	T	T		N	N	D			94.30	0.77	0.82	1.05	1.09	0.73	0.84	0.67	0.01	-0.06	0	2	6	4
165	-	W	W		F	W				0.00	0.64	0.73	1.02	0.98	0.66	0.77	0.61	0.00	-0.08	10	14	7	7
166	-	S	G	-0.46	S	S				85.29	0.78	0.96	1.17	1.18	0.83	0.92	0.81	0.74	-0.06	0	2	8	6
167	-	S	S		S	S				60.17	0.76	0.90	1.01	1.15	0.80	0.88	0.80	0.28	0.11	0	2	4	4
168	-	V	V		I	V				1.65	0.57	0.68	0.72	0.86	0.60	0.65	0.63	0.00	0.19	0	4	7	7
169	-	R	R		T	R				150.90	0.50	0.60	0.64	0.77	0.53	0.57	0.59	1.35	0.37	0	4	4	4
170	-	V	V		V	V				0.21	0.39	0.47	0.47	0.60	0.40	0.44	0.44	0.00	0.45	0	0	9	8
171	-	P	P		P	P				23.76	0.37	0.43	0.42	0.54	0.35	0.39	0.38	0.04	0.69	0	0	6	6
172	B	T	T		T	T				0.00	0.36	0.40	0.39	0.49	0.34	0.36	0.35	0.00	0.66	0	0	8	7
173	B	L	L		L	L				0.11	0.36	0.39	0.39	0.47	0.34	0.36	0.33	0.00	0.48	0	0	7	6
174	B	I	I		I	I				0.00	0.39	0.40	0.42	0.51	0.36	0.41	0.34	0.00	0.46	0	0	6	6
175	B	I	I		F	I		1.00	0.02	0.47	0.46	0.49	0.59	0.41	0.47	0.37	0.00	0.14	0	0	8	6	
176	B	G	G		A	G			0.28	0.55	0.53	0.52	0.62	0.46	0.49	0.41	0.00	-0.03	0	0	7	6	
177	B	A	A		C	A		-1.00	0.43	0.63	0.67	0.59	0.88	0.55	0.57	0.48	0.00	-0.03	0	0	9	8	
178	-	E	E		E	E			65.80	0.73	0.86	0.74	1.33	0.82	0.66	0.71	0.62	0.02	0	0	8	8	
179	-	N	N		N	L			84.33	0.73	0.89	0.81	1.69	0.93	0.71	0.83	0.30	-0.19	99	0	7	5	
180	-	D	D		D	D			7.25	0.76	0.96	0.75	1.66	0.85	0.69	0.71	0.00	-0.42	99	142	10	10	
181	-	T	T		S	T		-1.00	126.15	0.85	1.15	0.93	1.78	1.06	0.85	0.83	0.18	-0.18	99	0	5	4	
182	-	I	I		I	I			104.44	0.82	0.92	0.87	1.16	0.82	0.76	0.71	-0.29	-0.32	99	142	5	5	
183	-	A	A		A	A			1.12	0.89	0.70	0.86	0.90	0.66	0.65	0.58	0.00	-0.19	99	142	6	4	
184	-	S	S		P	P			37.81	0.76	0.74	0.82	0.90	0.67	0.59	0.65	-0.05	-0.16	12	32	9	8	
185	-	V	V		V	V			19.13	0.67	0.68	0.77	0.83	0.55	0.59	0.66	0.01	0.01	0	0	8	8	
186	H	R	R		S	A			200.23	0.74	0.68	0.77	0.82	0.56	0.60	0.69	0.28	-0.13	0	0	6	6	
187	H	S	S		S	T			56.50	0.71	0.67	0.77	0.81	0.62	0.56	0.66	0.47	-0.37	12	32	5	5	
188	H	H	H		H	H		1.00	17.95	0.55	0.50	0.57	0.65	0.48	0.46	0.51	0.05	-0.17	0	0	9	9	
189	H	S	S		A	A			0.02	0.56	0.49	0.51	0.61	0.47	0.48	0.51	0.00	-0.01	12	32	7	6	
190	H	E	E		S	K			38.45	0.65	0.54	0.55	0.64	0.50	0.53	0.55	0.11	-0.33	0	0	8	7	
191	H	P	P		P	P			36.58	0.63	0.56	0.56	0.66	0.52	0.56	0.50	0.05	-0.23	0	1	7	6	
192	H	F	F		F	F			0.00	0.54	0.50	0.51	0.62	0.47	0.50	0.44	0.00	-0.04	12	32	9	8	
193	H	Y	Y		Y	Y			18.08	0.56	0.52	0.53	0.63	0.48	0.52	0.49	0.02	0.04	0	0	7	7	
194	H	N	N		N	N			95.69	0.63	0.59	0.62	0.72	0.54	0.59	0.56	0.24	0.01	0	1	6	6	

195	H	S	S		S	S				-1.00	26.24	0.61	0.61	0.67	0.78	0.55	0.63	0.55	0.09	0.13	0	1	8	6	
196	H	L	L		I	L					5.73	0.59	0.63	0.68	0.82	0.56	0.66	0.62	0.00	0.32	0	0	7	7	
197	-	P	P		P	P					35.31	0.70	0.75	0.82	1.04	0.69	0.82	0.87	0.10	-0.08	0	4	6	6	
198	-	G	G		S	S	S				59.15	0.91	0.98	1.00	1.39	0.96	1.00	1.07	0.41	-0.31	0	0	4	4	
199	-	S	S		T	S					117.24	0.94	1.05	1.03	1.48	0.97	1.03	1.07	0.73	-0.17	0	4	4	4	
200	-	L	L		T	I					18.69	0.64	0.74	0.74	1.27	0.63	0.72	0.75	0.08	-0.01	0	4	6	6	
201	-	D	D		A	S					46.06	0.54	0.59	0.58	0.80	0.49	0.60	0.52	0.14	-0.01	0	78	6	5	
202	B	K	K		K	K					18.81	0.56	0.59	0.58	0.70	0.53	0.61	0.46	0.02	-0.13	0	78	9	7	
203	B	A	A		A	A					0.00	0.45	0.49	0.48	0.58	0.43	0.47	0.42	0.00	0.09	0	0	6	6	
204	B	Y	Y		F	Y			1.00		0.08	0.47	0.49	0.49	0.57	0.43	0.45	0.42	0.00	-0.08	0	0	8	7	
205	B	L	L		F	L					0.00	0.47	0.49	0.50	0.55	0.42	0.43	0.40	0.00	-0.30	0	0	6	5	
206	B	E	E		E	E					10.43	0.59	0.60	0.60	0.64	0.51	0.46	0.41	0.00	-0.10	0	0	9	8	
207	B	L	L		I	L					1.27	0.72	0.74	0.71	0.81	0.61	0.54	0.45	0.00	-0.03	4	32	8	7	
208	-	D	D		N	D	N		1.00		61.28	0.91	1.01	0.86	1.28	0.79	0.71	0.53	0.61	-0.14	4	32	8	8	
209	-	G	G		G	G				-1.00	46.72	1.04	1.15	0.97	2.08	0.92	0.88	0.67	0.94	-0.07	0	32	5	4	
210	-	A	A		G	A					8.22	0.87	1.34	0.83	1.88	1.09	0.73	0.63	0.00	-0.23	0	32	9	6	
211	-	S	S		S	T					48.60	0.84	1.14	0.87	2.24	1.00	0.69	0.70	0.00	-0.11	99	142	8	7	
212	-	H	H		H	H					8.93	0.88	1.36	1.03	2.23	0.89	0.63	0.57	0.00	-0.41	0	0	5	5	
213	-	F	F		S	F			1.00	-1.00	111.23	1.10	1.38	1.05	1.93	0.86	0.73	0.59	-0.16	-0.05	99	142	6	6	
214	-	A	A		C	A				-1.00	2.37	1.17	1.33	1.51	2.00	0.94	0.74	0.75	0.00	-0.02	0	142	9	9	
215	-	P	P		A	P				-1.00	0.00	1.19	1.30	1.61	2.95	0.91	0.86	0.94	0.00	-0.16	99	142	5	4	
216	-	N	N		N	N					40.22	1.22	1.40	1.41	3.00	0.77	1.04	1.02	-0.10	-0.15	99	142	7	5	
217	-	L	L		G	I					88.44	1.26	1.93	1.61	2.59	0.81	1.33	1.24	-0.02	-0.07	0	142	6	4	
218	-	S	S		G	P					98.41	1.44	1.87	1.74	2.35	0.75	1.88	1.64	0.09	-0.29	1	142	4	7	
219	-	N	N		N	N				-1.00	-1.00	69.48	1.49	1.51	1.61	1.84	0.66	1.41	1.49	0.02	-0.35	1	0	6	5
220	-	T	T		S	K					93.77	1.17	1.46	1.29	1.56	0.66	1.27	1.37	0.03	-0.33	4	32	5	5	
221	-	S	-		N	-	Y	T			19.29	0.85	0.85	0.85	1.19		0.85	0.80	0.06	0.02	4	32	7	9	
222	H	S	-		G	-		S			2.24	0.77	0.83	0.73	1.09		0.89	0.74	0.00	-0.14	1	142	8	9	
223	H	G	-		G	-		S			20.27	0.72	0.82	0.68	0.97		0.80	0.66	-0.10	-0.21	1	4	7	8	
224	H	T	T		L	I					12.86	0.63	0.68	0.60	0.91	0.60	0.58	0.50	0.01	0.06	4	32	7	7	
225	H	I	I		L	I					0.00	0.56	0.56	0.54	0.79	0.51	0.53	0.44	0.00	0.05	0	142	8	8	

226	H	A	A		G	G				0.00	0.53	0.59	0.53	0.70	0.48	0.57	0.49	0.00	0.28	0	0	8	8
227	H	K	K		K	K				27.81	0.52	0.62	0.54	0.70	0.47	0.55	0.47	-0.12	0.28	0	24	9	10
228	H	Y	Y		Y	Y				5.53	0.49	0.55	0.48	0.67	0.45	0.47	0.42	0.00	0.46	0	2	8	7
229	H	S	S		G	S				0.00	0.46	0.49	0.46	0.60	0.43	0.46	0.42	0.00	0.63	0	0	8	7
230	H	I	I		V	I				0.00	0.44	0.49	0.45	0.57	0.40	0.48	0.39	0.00	0.32	45	73	9	7
231	H	S	S		S	A				0.00	0.43	0.49	0.43	0.61	0.39	0.45	0.38	0.00	0.19	0	24	9	9
232	H	W	W		W	W				0.81	0.40	0.45	0.40	0.55	0.37	0.40	0.36	0.00	0.49	0	78	8	7
233	H	L	L		M	L				0.02	0.40	0.45	0.42	0.54	0.37	0.42	0.35	0.00	0.54	0	0	8	8
234	H	K	K		K	K				3.07	0.42	0.48	0.43	0.62	0.38	0.45	0.37	0.00	0.37	45	73	10	9
235	H	R	R		R	R				37.63	0.43	0.51	0.43	0.64	0.39	0.44	0.39	-0.03	-0.04	0	78	11	10
236	H	F	F		F	F				24.14	0.45	0.53	0.50	0.65	0.46	0.48	0.43	0.05	-0.21	0	78	8	7
237	H	V	V		L	V				2.88	0.41	0.51	0.50	0.66	0.44	0.47	0.42	0.00	-0.35	0	0	8	9
238	-	D	D		D	D				0.06	0.44	0.56	0.54	0.78	0.45	0.51	0.46	0.00	-0.64	0	4	8	8
239	-	D	D		N	N				74.77	0.52	0.63	0.63	0.87	0.48	0.53	0.52	0.87	-0.40	0	4	6	6
240	-	D	D		D	D				0.06	0.55	0.65	0.66	0.95	0.47	0.57	0.58	0.00	-0.28	0	73	8	8
241	H	T	T		T	T				86.00	0.72	0.80	0.84	1.24	0.60	0.65	0.79	0.94	-0.62	0	4	6	5
242	H	R	R		R	R				53.04	0.71	0.84	0.84	1.33	0.61	0.66	0.87	0.27	-0.60	0	73	5	6
243	H	Y	Y		Y	Y				9.78	0.64	0.73	0.71	1.19	0.57	0.60	0.67	0.00	-0.24	45	73	8	8
244	H	T	T		S	T				34.64	0.87	0.80	0.83	1.29	0.69	0.69	0.77	-0.19	0.03	0	78	8	9
245	H	Q	Q		Q	Q		-1.00		108.41	1.13	0.86	1.01	1.30	0.88	0.88	0.75	-0.70	0.20	0	0	6	7
246	H	F	F		F	F				17.82	1.27	0.82	0.95	1.17	0.98	0.94	0.68	-0.04	0.16	0	24	7	6
247	H	L	L		L	L				0.28	1.12	0.84	0.86	1.18	0.87	0.89	0.88	0.00	0.26	0	78	7	8
248	-	C	C		C	C				42.42	1.07	1.07	1.01	1.56	0.90	1.04	1.21	-0.09	0.10	0	0	8	7
249	-	P	P		G	P				123.56	1.38	1.38	1.41	1.95	1.30	1.47	1.60	-0.14	0.10	0	78	4	4
250	-	G	G		P	G				20.36	1.47	1.57	1.62	1.98	1.34	1.57	2.12	-0.01	0.17	0	78	6	5
251	-	P	P		N	P				27.67	1.58	1.85	2.26	2.31	1.20	1.62	2.19	0.05	0.13	0	0	3	5
252	-	S	S		H	R				87.74	2.25	2.75	2.99	2.91	1.41	2.00	3.10	0.27	0.17	0	0	3	3
253	-	T	T		E	D				68.91	3.26	3.50	3.91	4.05	1.79	2.47	2.92	-0.08	0.07	0	2	7	5
254	-	G	G		A	G				32.00	3.38	3.54	4.61	4.78	2.48	2.79	3.15	-0.10	-0.10	0	0	5	6
255	-	L	W	0.08	D	L				163.57	3.66	3.29	4.74	4.91	2.64	2.60	3.80	-0.02	0.14	0	0	4	6
256	-	F	G	3.43	L	F				213.71	3.26	2.70	3.90	4.28	2.24	2.11	3.35	-0.62	0.02	0	0	3	6

257	-	S	S		T	G					14.98	1.91	1.69	2.29	2.76	1.50	1.56	2.17	0.00	-0.10	0	32	7	6
258	-	D	D		R	E					76.81	1.06	1.20	1.14	1.47	0.99	0.92	0.83	0.12	0.04	4	32	4	4
259	B	V	V		I	V					3.20	0.83	0.85	0.85	1.06	0.75	0.73	0.58	0.00	-0.03	0	0	9	10
260	B	E	E		S	E					65.18	0.80	0.80	0.79	1.00	0.72	0.69	0.55	0.80	-0.22	0	0	8	7
261	B	E	E		E	E					44.00	0.71	0.73	0.72	0.83	0.63	0.63	0.56	0.41	-0.43	0	0	7	5
262	B	Y	Y		Y	Y					32.34	0.65	0.69	0.67	0.77	0.56	0.61	0.58	0.09	-0.33	0	78	5	6
263	B	R	R		R	R					80.92	0.63	0.66	0.71	0.73	0.62	0.68	0.53	0.29	-0.17	0	0	5	5
264	-	S	S		G	S					35.32	0.75	0.89	0.98	0.92	0.84	0.88	0.54	0.11	-0.46	0	78	6	6
265	-	T	T		N	T					53.35	1.06	1.30	1.38	1.31	1.23	1.36	0.63	-0.03	-0.46	0	78	5	5
266	-	C	C		C	C					48.49	1.33	1.52	1.88	1.56	1.39	1.63	1.14	0.07	-0.46	0	0	5	4
267	-	P	P		P	P					41.69	2.14	2.52	3.57	2.54	1.86	2.15	2.17	-0.12	-0.46	0	78	3	2
268		-	F		Y	F					-													