# Supporting Information for

***Characterization and engineering of a plastic-degrading aromatic polyesterase***

Harry P. Austin[a,1], Mark D. Allen[a,1], Bryon S. Donohoe[b,1], Nicholas A. Rorrer[c,1], Fiona L. Kearns[d,1], Rodrigo L. Silveira[b,e], Benjamin C. Pollard[d], Graham Dominick[c], Ramona Duman[f], Kamel El Omari[f], Vitaliy Mykhaylyk[f], Armin Wagner[f], William E. Michener[c], Antonella Amore[b], Munir S. Skaf[e], Michael F. Crowley[b], Alan W. Thorne[a], Christopher W. Johnson[c], H. Lee Woodcock[d]*, John E. McGeehan[a]*, and Gregg T. Beckham[c]*

[a]Molecular Biophysics Laboratories, School of Biological Sciences, Institute of Biological and Biomedical Sciences, University of Portsmouth, United Kingdom
[b]Biosciences Center, National Renewable Energy Laboratory, Golden CO, USA
[c]National Bioenergy Center, National Renewable Energy Laboratory, Golden CO, USA
[d]Department of Chemistry, University of South Florida, 4202 E. Fowler Ave., CHE205, Tampa, Florida 33620-5250, United States
[e]Institute of Chemistry, University of Campinas, Campinas, Sao Paulo, 13083-970, Brazil
[f]Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, United Kingdom

* Email: gregg.beckham@nrel.gov, john.mcgeehan@port.ac.uk, hlw@mail.usf.edu
[1] These authors contributed equally to this work

**SI Materials and Methods**

**Cloning, expression, purification and crystallization.** The gene encoding PETase from *Ideonella sakaiensis* strain 201-F6 (Genbank GAP38373.1) was codon optimized for expression in Escherichia coli using the OPTIMIZER codon optimization software (1) (guided random method) based on the codon usage of highly expressed genes in E. coli K-12 MG1655 (2). The stop codon was omitted, overlaps were added for assembly into the expression vector pET-21b(+) (Millipore Sigma), and one of the six CAC His codons was changed to CAT to enable DNA synthesis. This 927 bp sequence (**Fig. S2B**) was synthesized by Integrated DNA Technologies as a gBlock DNA fragment and assembled using NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs) into pET-21b(+) digested with NdeI and XhoI, generating a C-terminal His-Tag. This plasmid was modified using the Q5® Site-Directed Mutagenesis Kit (New England Biolabs) to generate plasmids for expression of the PETase containing the S238F/W159H or W185A mutations. For the S238F/W159H double mutant, the PETase expression plasmid was amplified by PCR using the Q5 polymerase with primers oCJ699 (5'-CACTCAATGGGGGGCGGC-3') and oCJ700 (5'-GCCCATCACACCCATGCG-3'), which introduce the W159H mutation (underlined), and the product was treated with the Kinase, Ligase, and DpnI (KLD) Enzyme Mix according to the manufacturer's instructions. The S238F mutation was subsequently introduced in a second round of site-directed mutagenesis with primers oCJ703 (5'-TTCTGTGCCAACTCTGGGAACAGC-3') and oCJ704 (5'-GTGGCTACCGCCGTTAATTTCCAG-3'). The W185A mutation was introduced by site-directed mutagenesis with primers oCJ701 (5'-GACTCTTCAACCAACTTCAGCAGTGTTAC-3') and oCJ702 (5'-CGCTGGCGCCTGCGGTG-3'). The sequence of the resulting plasmids was confirmed using Sanger sequencing performed by GENEWIZ, Inc.

The protein was expressed in E. coli C41(DE3) in 2xTY media and induced by the addition of 1 mM final concentration IPTG. The cells were harvested by centrifugation, disrupted by sonication, and the cell debris removed by centrifugation at 20,000 rpm for 30 min. The protein was purified by Ni-affinity chromatography followed by gel filtration on a Superdex 75 HR column. The protein was concentrated to 10 mg/mL and dialyzed into 10 mM Tris, pH 7.5, 100 mM NaCl for crystallography. PETase was crystallized at 10 mg/mL by sitting-drop vapor diffusion. Several conditions yielded crystals with 5 conditions being used to obtain datasets (**Table S1**).

**Structure solution and analysis.** The crystals were flash-frozen in liquid nitrogen after the addition of glycerol to 20% while leaving the other components of the mother liquor at the same concentration. PETase crystals belonging to space group $C222_1$ were used to obtain phase information using the I23 long-wavelength beamline at Diamond. Native sulfur-SAD data were obtained from 3,600 images collected at $\lambda$ = 2.455 Å with 0.1° increments on a curved Pilatus 12M detector. All images were integrated using XDS (3) and scaled using SCALA (4). Phases were obtained using PHASERSAD (5) in the CCP4i software in combination with PARROT and SHELXD (6). The initial output was subsequently built using BUCCANEER and further refined using iterative rounds of COOT (7) and PHENIX (8). Three molecules of PETase were observed in the asymmetric unit of the $C222_1$ S-SAD dataset (6EQD). A native $P2_12_12_1$ dataset was collected for PETase using the I23 beamline at 1.2398 Å with 1,800 images collected at 0.1° increments, and a further 800 images collected using an adjustment of the kappa goniometer to increase the completion of the dataset from 93.8% to 99.8% in the outer shell (99.9% overall) (6EQE). Three additional datasets were collected at beamline I03 of the Diamond Light Source, with three independent PETase molecules present in 6EQG and 6EQH and one in 6EQF. The structures of



www.pnas.org/cgi/doi/10.1073/pnas.1718804115

PETase obtained in other conditions and space groups were obtained using molecular replacement from a refined molecule of PETase obtained initially from the sulfur-SAD data. All structures were refined using iterative rounds of COOT (7) and PHENIX (9). Cell constants, crystallographic data, and details of the refined models are shown in **Table S1**. Figures were generated with PyMOL (Schrödinger, LLC). Sequence analysis was performed using Clustal Omega (10) and alignments rendered with ESPript (11).

**Protein differential scanning calorimetry.** The $T_m$ of PETase was measured using a Microcal model VP-DSC calorimeter (Microcal, Inc., Northampton, USA). Lyophilized protein was re-suspended in water, and then dialyzed against 80 mM phosphate buffer, pH 8 by means of centrifugal filter units (Sartorius, AG, Germany). For thermal stability analyses, protein samples were diluted up to 50 and 100 µg/mL in 80 mM phosphate buffer, pH 8. The calorimeter scan rate was 60°C/h over a range of 25 to 80°C. Experiments were performed in triplicate and data analysis was completed by Origin for DSC software (OriginLab, Northampton, USA).

**Molecular dynamics.** The MD simulations of *I. sakaiensis* PETase started from the 0.92 Å resolution structure reported in this work, and simulations of *T. fusca* cutinase started from the crystal structure reported by Roth *et al*. (12) (PDB: 4CG1). Protonation states of titratable residues were estimated using H++ (13) at pH 7.0. For PETase, all Asp and Glu are deprotonated, and the His residues are neutral with a single proton on the Nδ nitrogen. For the cutinase, Glu253 is the only protonated residue among all Asp and Glu, and all the His residues are neutral, with His106 and His208 protonated on the Nδ nitrogen, and His77, His129, His156, and His184 protonated on the Nε nitrogen. As Asp246 is missing in the cutinase structure, this residue was taken from another T. fusca cutinase structure of lower resolution (PDB 4CG3) (12). The systems were then immersed in a rectangular box of water at least 15 Å thick containing 0.1 M NaCl plus excess counterions to neutralize the systems. The final systems comprised approximately 46,000 atoms.

The CHARMM36 force field (14) along with the TIP3P water model (15) was used to model the interactions, and all the MD simulations were performed with NAMD2.12 (16). The temperature and pressure were kept constant at 300 K and 1 bar, using the Langevin thermostat and barostat, respectively, as implemented in NAMD. Chemical bonds involving hydrogen atoms were constrained at their equilibrium values, and a time step of 2 fs was employed to integrate the equation of motions. Periodic boundary conditions were employed. Short-range interactions were truncated at a cutoff radius of 12 Å and long-range interactions were computed with particle mesh Ewald (17). The simulations were conducted according to the following steps: (1) with all the protein atoms fixed, 1,000 steps of energy minimization followed by 100 ps of MD simulation; (2) with all the protein alpha carbons fixed, 2,000 steps of energy minimization followed by 400 ps of MD simulation; (3) with all atom free, 2,000 steps of energy minimization followed by 10 ns of MD. After steps 1-3, which were considered as equilibration and not considered for analyses, simulations of 200 ns were run. Three independent simulations, starting from (1), were performed for each system.

**Substrate docking.** PETase Structure Preparation: The PETase protein structure was prepared with Schrödinger's Protein Preparation Wizard (18-20). Hydrogen bonds were optimized with PropKa (at pH 7.0), and a restrained minimization was conducted using the OPLS3 force field (21) until protein heavy atoms converged to RMSD of 0.30 Å from initial crystal structure positions. A $Cl^-$ ion was crystallized in the binding site and interacts with Ser160 in the binding site. However, we know that this $Cl^-$ quickly moves to the bulk solvent during MD simulations with explicit water, therefore, we determined that deleting it for the purposes of predicting PET binding modes was an acceptable solution.

**Ligand structure preparation.** PET tetramer structures (capped at both ends with ethanol groups) were prepared and energetically minimized using the Schrödinger LigPrep module according to the OPLS3 force field (21). Ionizations states of the dimer and trimer were requested with Epik at target pH of 7.0 +/- 2.0 (22-25) although no additional ionization states were generated.

**Induced fit docking (IFD)**. IFD is Schrödinger's flexible ligand/flexible receptor docking protocol (24-26). To virtually model protein flexibility that might be seen upon ligand association, IFD has a mutation/back-mutation protocol in which selected binding site residues are mutated first to alanine, which "opens up" the binding site, a ligand conformer is initially docked into the binding site according to the Glide SP scoring function (27-29), Prime (side chain prediction tool) (30, 31) then sequentially back-mutates residues to their original residue identity with loose minimization routines built in to alleviate bad atom contacts and confer additional active site flexibility. Finally, the ligand is then re-docked into the new binding site configuration using the Glide XP scoring function. In this current work, the center of the binding site was defined as the geometric center of catalytic triad residues (Ser160, Asp206, His237). The OPLS3 force field was used for both the protein and ligand. Residues selected for the mutation/back-mutation routine were His237, Asp206, Ser160, Ile208, Trp159/H159, Trp185, Met161, Tyr87, F238. Some of these residues are hypothesized to be important for recognition and binding of PET and PEF to PETase (Typ87, Ile208, Trp185, Trp159), thus we wanted to

ensure these residues were fully flexible to capture any large side chain movements necessary for binding.

**Polymer polycondensation.** Unless otherwise noted, all chemicals were obtained and used as is from Sigma Aldrich.

All polymers, except for polylactic acid, were synthesized via a polycondensation reaction. All polymerization reactions were conducted in a three-necked round bottom flask fixed with a top stirring mechanical motor, a Dean-stark condensation trap, and nitrogen flow lines attached. Initially, the reactor was charged with 1 mol of diacid or diacyl chloride (terephthalic acid, furan-2,5-dicarbonyl chloride, or succinic acid) to 1.1 moles of diol (ethylene glycol or butanediol) with 0.1 wt% titanium butoxide as a transesterification catalyst. Furan-2,5-dicarbonyl chloride was synthesized from FDCA by adding phosphoryl chloride at a ratio of 3.3:2 (phosphoryl chloride:FDCA) into a solution of DCM and allowing the solution to stir overnight. The solid was then vacuum filtered from solution and used without further purification. Initially, the reactor was heated up to 180°C and allowed to stir under nitrogen for two hours. Following the initial pre-polymerization, vacuum was applied to the system, and in the case of PET and PEF, the temperature was slowly ramped to 220°C. This polymerization step was allowed to continue for six hours, after which heat and stirring to the reactor was terminated and the reactor was allowed to cool to room temperature. Following this, the polymers were dissolved in a solvent (triflouroacetic acid for PET and PEF and chloroform for PBS) and then precipitated by adding the solution drop-wise to cold methanol. The polymer precipitate was then filtered and subsequently allowed to dry in a vacuum oven for 24 hours.

**Ring opening polymerization.** PLA was prepared via ring opening polymerization using stannous octoate in the same reactor as described above. Initially, the reactor was heated to 80°C under vacuum to remove any excess water for one hour. Following water removal, 0.1% w/v stannous octoate was charged into the reactor and the temperature was elevated to 140°C for two hours. After two hours, the temperature was slowly ramped up to 180°C and allowed to polymerize for six more hours, after which heat and stirring to the reactor was terminated and the reactor was allowed to cool to room temperature. PLA was subsequently dissolved in chloroform, precipitated in methanol, filtered, and allowed to dry in a vacuum oven for 24 hours prior to use.

**Polymer sample preparation.** All polymers were dissolved in their respective solvent, as described above and the solvent was allowed to evaporate to produce a film for each polymer. The films were then allowed to anneal above their glass transition temperature overnight in a vacuum oven. To prepare the discs for enzymatic testing, coupons were cut using a biopsy needle of diameter 5 mm. Due to their rigid nature, some of the polymers fractured upon cutting and were not perfect circles. Additionally, a sample from a commercial soft drink bottle was prepared by directly cutting a coupon from a cleaned plastic bottle.

**Thermal measurements for polymer samples.** All thermal measurements were conducted on a TA Instruments Q-500 Digital Scanning Calorimeter (DSC). In order to probe the native film with introducing additional thermal history, the first scan of the polymers is reported. Samples were placed in aluminum hermetically-sealed pans. Scans were conducted at a rate of 10°C/min from 60°C below the literature glass transition temperature to 30°C above the literature melting temperature for each respective polymer. Enthalpies of fusion were calculated by integration of the melting peak. Percent crystallinity is calculated by dividing the experimental heat of fusion for the polymer by the literature reported value for a 100% crystalline polymer, 140 J/g for PET, 137 J/g for PEF, 93/ J/g for PLA, and 200 J/g for PBS. In the case of PET, the cold crystallization peak was subtracted from the melting peak to report the final crystalinity. No other polymers exhibited cold crystallization.

**Structural characterization of polymers.** All polymer structures were determined by the use of a Bruker Avance III HD 400 MHz NMR Spectrometer with a 5mm BBO probe. Quantitative $^{1}H$ spectra were acquired with a 90° pulse of 14.5 µs and a 30 s recycle delay at room temperature. Deuterated trifluoroacetic acid (99.9% Cambridge Isotope Lab) with 1% w/w TMS was used as the solvent for PET and PEF while deuterated chloroform 99.9% Cambridge Isotope Lab) with 1% w/w TMS was used as a solvent for PLA and PBS.

**Water suppression NMR spectroscopy of digestion eluent.** To characterize the amount of products (BHET, MHET, TPA, FDCA) released during enzyme digestions, quantitative $^{1}H$ NMR spectra were obtained with the same previously mentioned spectrometer. Samples were run with the Bruker-native 1D WATERGRATE zggpwg Water Suppression pulse-program (which uses 90° water-selective pulses tuned to each sample) utilizing a relaxation delay, $d_1$, of 30 seconds and 64 scans. Succinic acid was used as an internal standard for peak integration. FDCA was identified as a 2H singular shift at ~7.5 ppm, and the entire peak was integrated to ascertain a quantitative value, which may have incorporated ethylene glycol esters. TPA, BHET, and MHET were all characterized by their aromatic shifts of a 4H singlet at 8.0 ppm, a 4H singlet at 8.1 ppm, and the two 2H aromatics doublets around 8.05 ppm respectively. Due to the limitations of the water suppression NMR method used, ethylene glycol or the ethylene glycol ester of FDCA were not quantitated.

**HPLC of PET digestion eluent.** To corroborate the quantitative NMR data, HPLC was also performed.

Standards of TPA and BHET were obtained from Sigma Aldrich and MHET was obtained from Acadechem, Co. Limited. Analyte analysis of samples was performed on an Agilent 1200 LC system (Agilent Technologies, Santa Clara, CA) equipped with a G1315A diode array detector (DAD). Each sample and standard was injected at a volume of 10 µL onto a Phenomenex Luna C18(2) column 5 µm, 4.6 x 150 mm column (Phenomenex, Torrance, CA). The column temperature was maintained at 30°C and the buffers used to separate the analytes of interest was 20 mM Phosphate buffer in water(A)/methanol(B). The separation was carried out using a gradient program of: (A) = 99% and (B) = 1% at time *t* = 0; (A) = 75% and (B) = 25% at time *t* = 15, (A) = 0% and (B) = 100% at *t* = 25 min; (A) = 0% and (B) = 100% at *t* = 26.00 min; (A) = 99% and (B) = 1% at *t* = 26.01 min; (A) = 99% and (B) = 1% at *t* = 35.00 min. The flow rate was held constant at 0.6 mL $min^{-1}$ resulting in a run time of 35 minutes. Calibration curve concentration for each analyte varied between the ranges of 5.0 – 100 µg $L^{-1}$. DAD wavelength of 240 nm was used for analysis of the analytes of interest. A minimum of 5-6 calibration levels was used with an $r^2$ coefficient of .995 or better for each analyte and a check calibration standard (CCS) was analyzed every 10 samples to insure the integrity of the initial calibration. HPLC data are in agreement with NMR data +/- 15 mg/L.

**SI References**


1. Puigbo P, Guzman E, Romeu A, & Garcia-Vallve S (2007) OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res. 35(suppl_2):W126-W131.
2. Puigbo P, Romeu A, & Garcia-Vallve S (2007) HEG-DB: a database of predicted highly expressed genes in prokaryotic complete genomes under translational selection. Nucleic Acids Res. 36(suppl_1):D524-D527.
3. Kabsch W (2010) XDS. Acta Cryst. D 66(2):125-132.
4. Collaborative CP (1994) The CCP4 suite: programs for protein crystallography. Acta Cryst. D 50(Pt 5):760.
5. McCoy AJ, et al. (2007) Phaser crystallographic software. J. Appl. Cryst. 40(4):658-674.
6. Sheldrick GM (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Cryst. D 66(4):479-485.
7. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Cryst. D 60(12):2126-2132.
8. Afonine PV, et al. (2012) Towards automated crystallographic structure refinement with phenix. refine. Acta Cryst. D 68(4):352-367.
9. Adams PD, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Cryst. D 66(2):213-221.
10. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7(1):539.
11. Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 42(W1):W320-W324.
12. Roth C, et al. (2014) Structural and functional studies on a thermostable polyethylene terephthalate degrading hydrolase from Thermobifida fusca. Appl. Microbiol. Biotechnol. 98(18):7815-7823.
13. Gordon JC, et al. (2005) H++: a server for estimating pKas and adding missing hydrogens to macromolecules. Nucleic Acids Research 33(suppl_2):W368-W371.
14. Best RB, et al. (2012) Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone φ, ψ and Side-Chain χ1 and χ2 Dihedral Angles. Journal of Chemical Theory and Computation 8(9):3257-3273.
15. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein ML (1983) Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics 79(2):926-935.
16. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. Journal of Computational Chemistry 26(16):1781-1802.
17. Darden T, York D, & Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics 98(12):10089-10092.
18. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, & Sherman W (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J. Comput. Aided Mol. Des. 27(3):221-234.
19. Anonymous (Schrödinger Release 2017-3: Schrödinger Suite 2017-3 Protein Preparation Wizard.
20. Anonymous (Schrödinger Release 2017-3: Impact, Schrödinger, LLC, New York, NY, 2016.
21. Harder E, et al. (2016) OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. J. Chem. Theory Comput. 12(1):281-296.
22. Greenwood JR, Calkins D, Sullivan AP, & Shelley JC (2010) Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. J. Comput. Aided Mol. Des. 24(6-7):591-604.
23. Shelley JC, et al. (2007) Epik: a software program for pKa prediction and protonation state generation for drug-like molecules. J. Comput. Aided Mol. Des. 21(12):681-691.
24. Farid R, Day T, Friesner RA, & Pearlstein RA (2006) New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. Bioinorg. Med. Chem. 14(9):3160-3173.
25. Sherman W, Day T, Jacobson MP, Friesner RA, & Farid R (2006) Novel procedure for modelling

ligand/receptor induced fit effects. J. Med. Chem. 49(2):534-553.
26. Sherman W, Beard HS, & Farid R (2006) Use of an induced fit receptor structure in virtual screening. Chem. Biol. Drug Des. 67(1):83-84.
27. Friesner RA, et al. (2006) Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein− ligand complexes. J. Med. Chem. 49(21):6177-6196.
28. Friesner RA, et al. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem. 47(7):1739-1749.
29. Halgren TA, et al. (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J. Med. Chem. 47(7):1750-1759.
30. Jacobson MP, et al. (2004) A hierarchical approach to all-atom protein loop prediction. Proteins: Struct. Funct. Bioinform. 55(2):351-367.
31. Jacobson MP, Friesner RA, Xiang Z, & Honig B (2002) On the role of the crystal environment in determining protein side-chain conformations. J. Mol. Biol. 320(3):597-608.

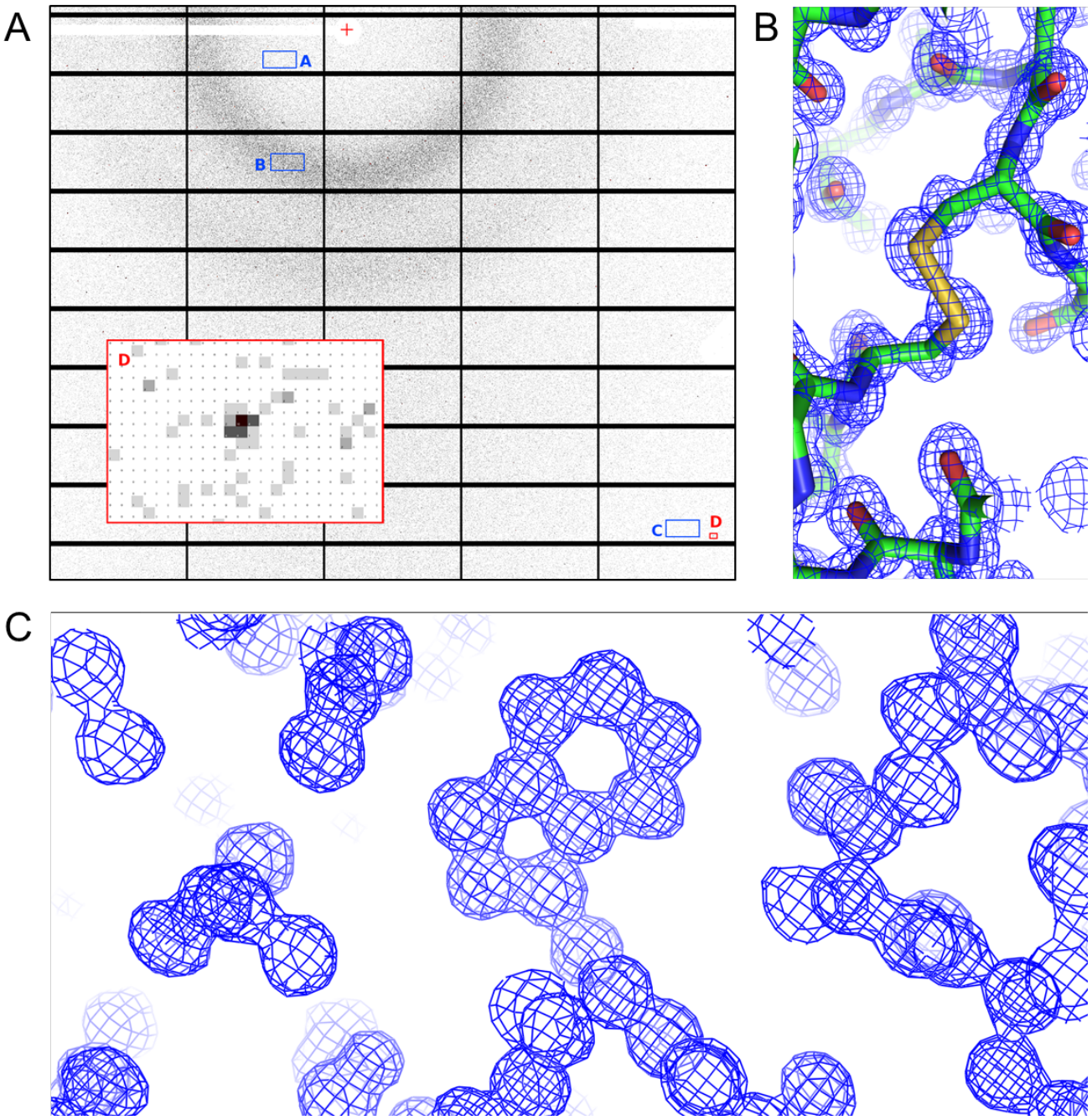
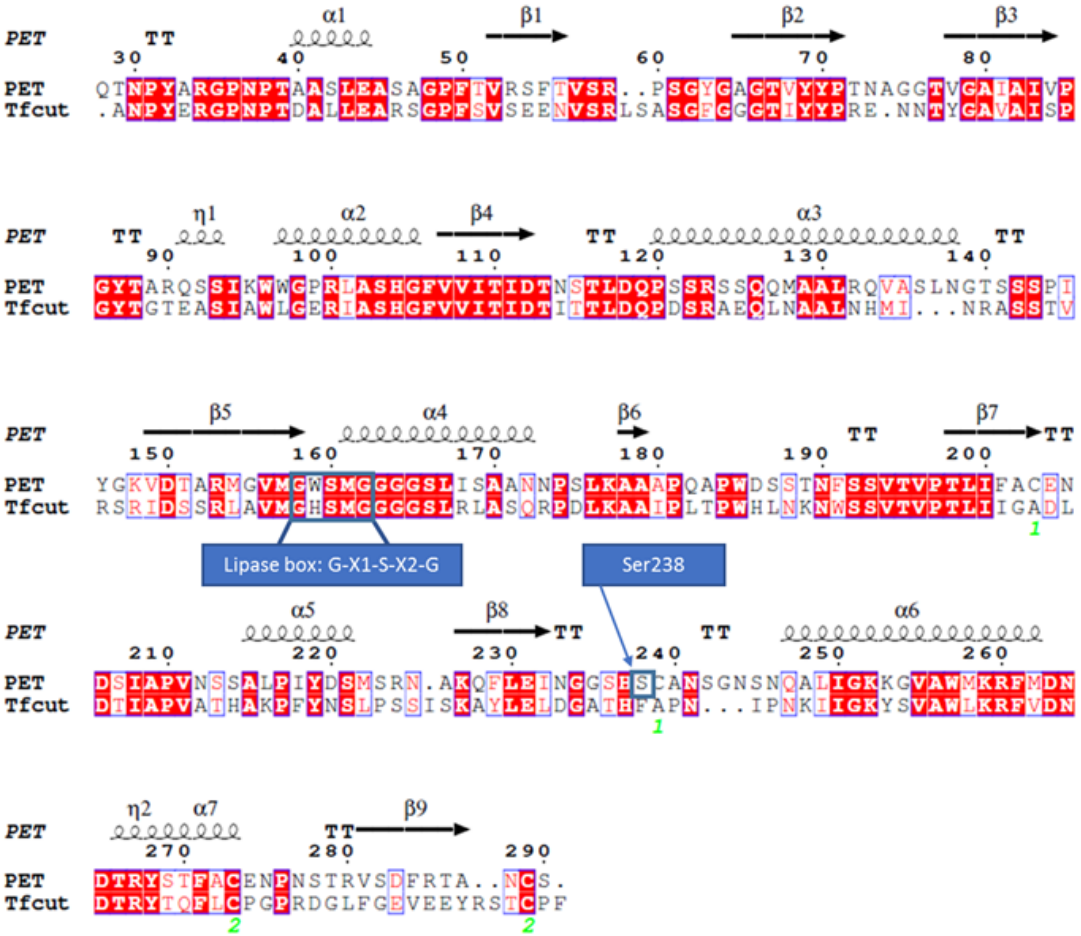



**Fig. S1.** High resolution X-ray crystallography data collection and analysis. (A) A representative section of a diffraction image from the 0.92 Å PETase structure determination. The mean counts in the blue boxes are A (low resolution): 0.3 counts, B (solvent ring): 1.6 counts and C (high resolution): 0.1 counts. The inset shows an enlarged portion around reflection -49 13 -11 at 1.0 Å resolution (box D) where each box represents an individual pixel. Values range from zero (white) and 1 (light grey) background counts, through to the center of this refection at 16 counts. (B) Representative electron density from the high-resolution structure (6EQE) showing continuous density across the disulfide bridge between Cys273 and Cys289. The $2F_o–F_c$ map is contoured at 1σ. (C) Representative electron density quality is shown centered around Trp257. The $2F_o–F_c$ map is contoured at 1σ.

A




B

ctttaagaaggagatataCAT**ATG**aacttcccccgtgcctcgcgccttatgcaggctgctgtgctgggcggccttatggccgtttccgcagcgg
ccaccgcgcagaccaatccgtatgcgcgcggccccaaccctaccgccgcctcgttggaagccagcgcgggaccctttaccgttcgtagcttta
ccgttagccgtccgtccggatatggtgcagggaccgtctattacccaaccaatgcaggcggcaccgttggcgcgattgcaatcgtccccgggta
caccgcgcgtcaaagcagcattaagtggtggggtccgcgcttagctagccatggctttgtggttattaccatcgatacgaacagcactctagacc
agcccagcagccgtagctcgcaacagatggccgcgcttcgtcaagttgcgagcttgaacgggaccagcagtagcccgatttacggaaaggtc
gatactgcccgcatgggtgtgatgggctggtcaatgggggggcggcggttcacttattagcgccgcgaacaacccgagtttaaaagcagcggca
ccgcaggcgccatgggactcttcaaccaacttcagcagtgttaccgtgccgacgctgattttcgcgtgcgagaatgatagcattgcaccggtgaa
cagcagcgcgctgccgatttatgatagcatgtcccgcaacgcaaaacagtttctggaaattaacggcggtagccactcttgtgccaactctggga
acagcaaccaggcactgatcggaaaaaaaggggttgcatggatgaaacgattcatggataatgacacccgttactcaaccttcgcctgtgaga
atcccaacagcacacgcgtgtcggattttcgcaccgcgaactgttccCTCGAGcaccaccatcaccaccac**tga**gatccggct

**Fig. S2.** PETase sequence analysis. (A) Sequence alignment of PETase (labelled PET, accession number: A0A0K8P6T7), against a PET-degrading cutinase from *T. fusca* (TFcut, accession number: AET05798). The signal sequences from both enzymes, as predicted by LipoP 1.0 Server, were excluded from the alignment for clarity. The squares shaded dark red indicate areas of identity with residues in red text indicating moderate conservation. The cartoons above the alignment denote secondary structure, with spirals representing alpha-helices, arrows representing beta-strands, and 'T' indicating turns. Differences in key residues around the active site are highlighted with a blue box, with residues 159 to 162 denoting the position of the conserved lipase box. The blue box spanning residues 237 – 238 denotes the presence of a serine in PETase next to the catalytic histidine, as compared to a phenylalanine in *T. fusca*. The green numbering below the sequences indicates the position of residues forming productive disulfide bridges as observed in the PETase crystal structure reported here. (B) Nucleotide sequence of the synthetic DNA fragment containing the gene encoding the PETase from *I. sakaiensis* optimized for expression in *E. coli*. Overlaps added for assembly into pET-21b(+) are underlined. The initiating ATG and stop codon of the His-tagged gene are in bold text. NdeI and XhoI restriction sites are capitalized.

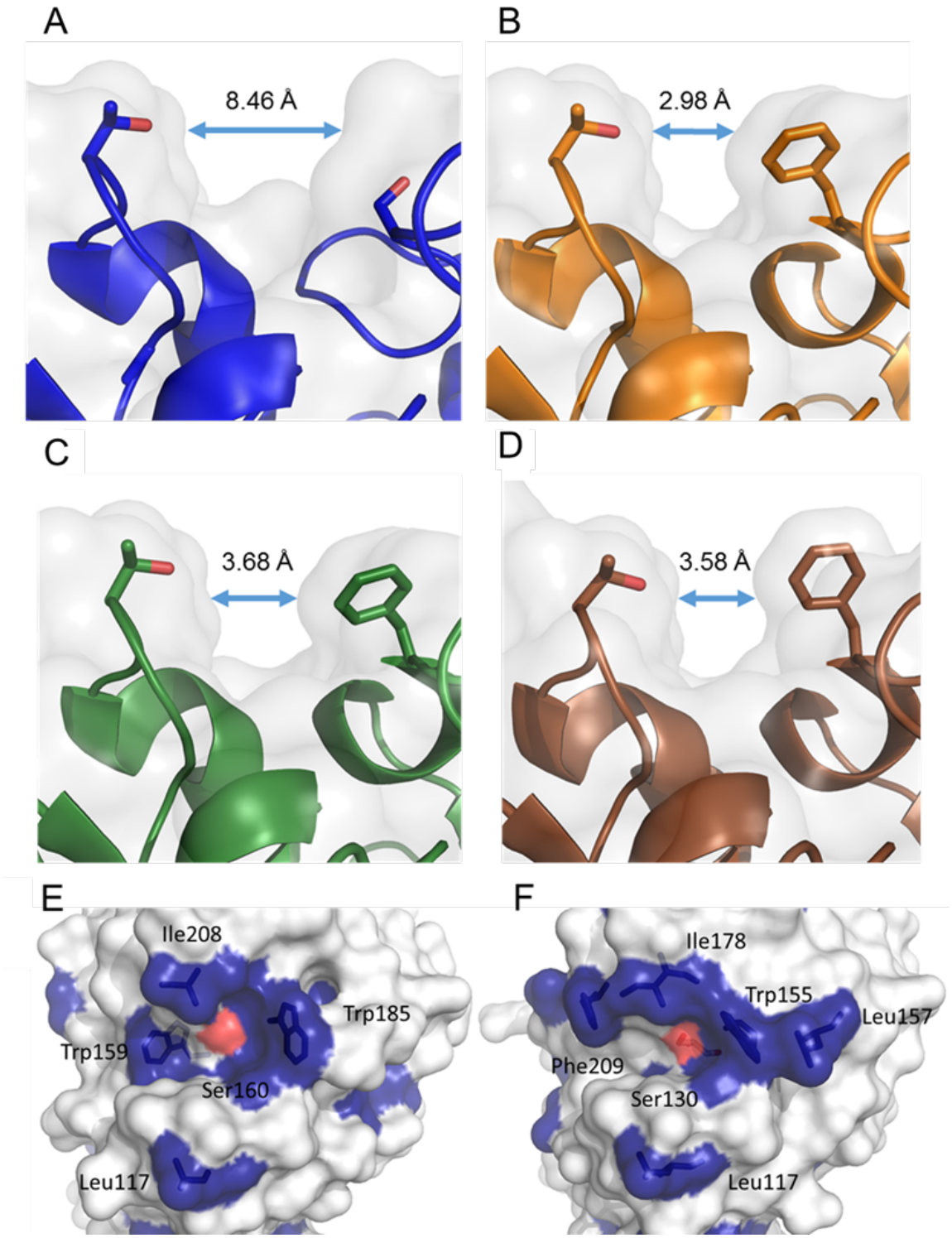



**Fig. S3.** Comparison of the active site cleft of PETase with cutinases. (A) PETase. Cleft width calculated from the distance between the van der Waals surface of T88 and S238 PDB ID: 6EQE (B) *T. fusca* cutinase with cleft width calculated from the distance between the van der Waals surface of T61 and F209 (PDB ID: 4CG1) (12) (C) *Thermobifida cellulosilytica* cutinase (PDB ID: code: 5LUI) with cleft width calculated from the distance between the van der Waals surface of T63 and F212 (D) leaf compost cutinase obtained from a metagenomics analysis of uncultured bacteria. Cleft width calculated from the distance between the van der Waals surface of T61 and F210 (PDB ID: 4EBO). Arrows with labels denote the width of the active site as determined from equivalent residues in each enzyme. (E) Hydrophobic adaptations in the active site of PETase with the catalytic residue, Ser160, colored in red. Residues in blue indicate the hydrophobic residues surrounding the active site triad. (F) A comparative view of the *T. fusca* hydrophobic distribution, with equivalent orientation and coloring.

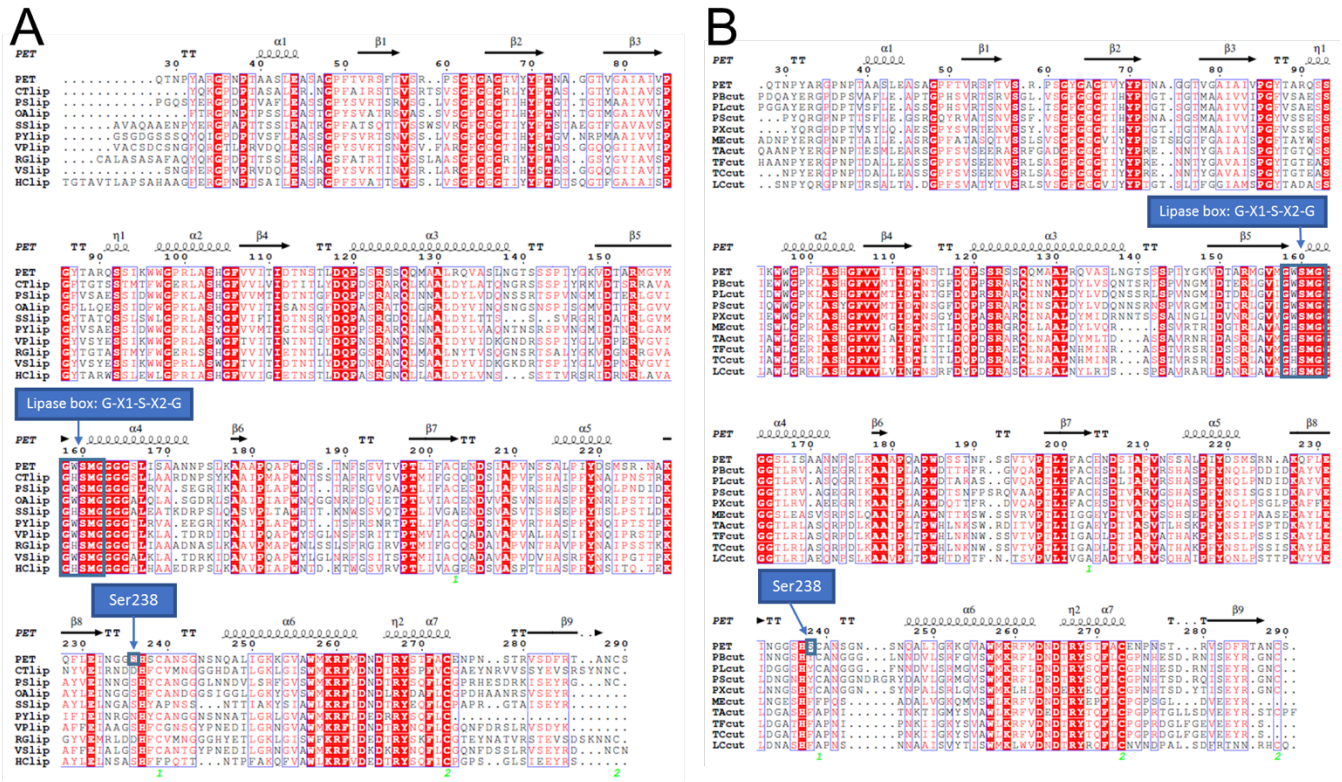



**Fig. S4.** Multiple sequence alignments of PETase with lipase and cutinase family members. (A) Multiple sequence alignment of PETase against members of the lipase family. Signal sequences, as predicted by LipoP 1.0 Server, were excluded from the alignment. As in Fig. S2A, the blue box spanning residues 159 to 162 highlights the conserved lipase box, while the box spanning 237-238 indicates the serine in PETase, which is occupied by a phenylalanine or tyrosine in most lipases. Aligned sequences with accession numbers are PET (PETase, *I. sakaiensis*, A0A0K8P6T7), CTlip (lipase, *Caldimonas taiwanensis*, WP_062195544), PSlip (lipase, *Pseudomonas saudimassiliensis*, CEA05385), OAlip (lipase, *Oleispira Antarctica*, CCK74972.1), SSlip (lipase, *Saccharothrix sp*, OKI36883.1), PYlip (triacylglycerol lipase, *Pseudomonas yangmingensis*, SFM35944), VPlip (lipase, *Vibrio palustris*, SJL84994), RGlip (lipase, (*Rhizobacter gummiphilus*, ARN20166), VSlip (Lipase, *Vibrio spartinae*, SIO95186) and HClip (lipase, *Herbidospora cretacea*, WP_061298849). (B) Multiple sequence alignment of PETase against members of the cutinase family with the same scheme. Aligned sequences are PET (PETase, *Ideonella sakaiensis*, A0A0K8P6T7), PBcut (cutinase, *Pseudomonas bauzanensis*, SER72431), Plcut (cutinase, *Pseudomonas litoralis*, SDS35700), PScut (cutinase, *Pseudomonas salegens*, SDU28434,), PXcut (cutinase, *Pseudomonas xinjiangensis*, SDS09569 ), MEcut (cutinase, *Micromonospora echinospora*, SCF30318), TAcut (cutinase, *Thermobifida alba*, ADV92525), TFcut (cutinase, *T. fusca*, AET05798), TCcut (cutinase, *Thermobifida cellulosilytica*, E9LVH8) and LCcut (cutinase, uncultured bacteria isolated from metagenomics analysis, AEV21261). Annotations are consistent with Fig. S2A.

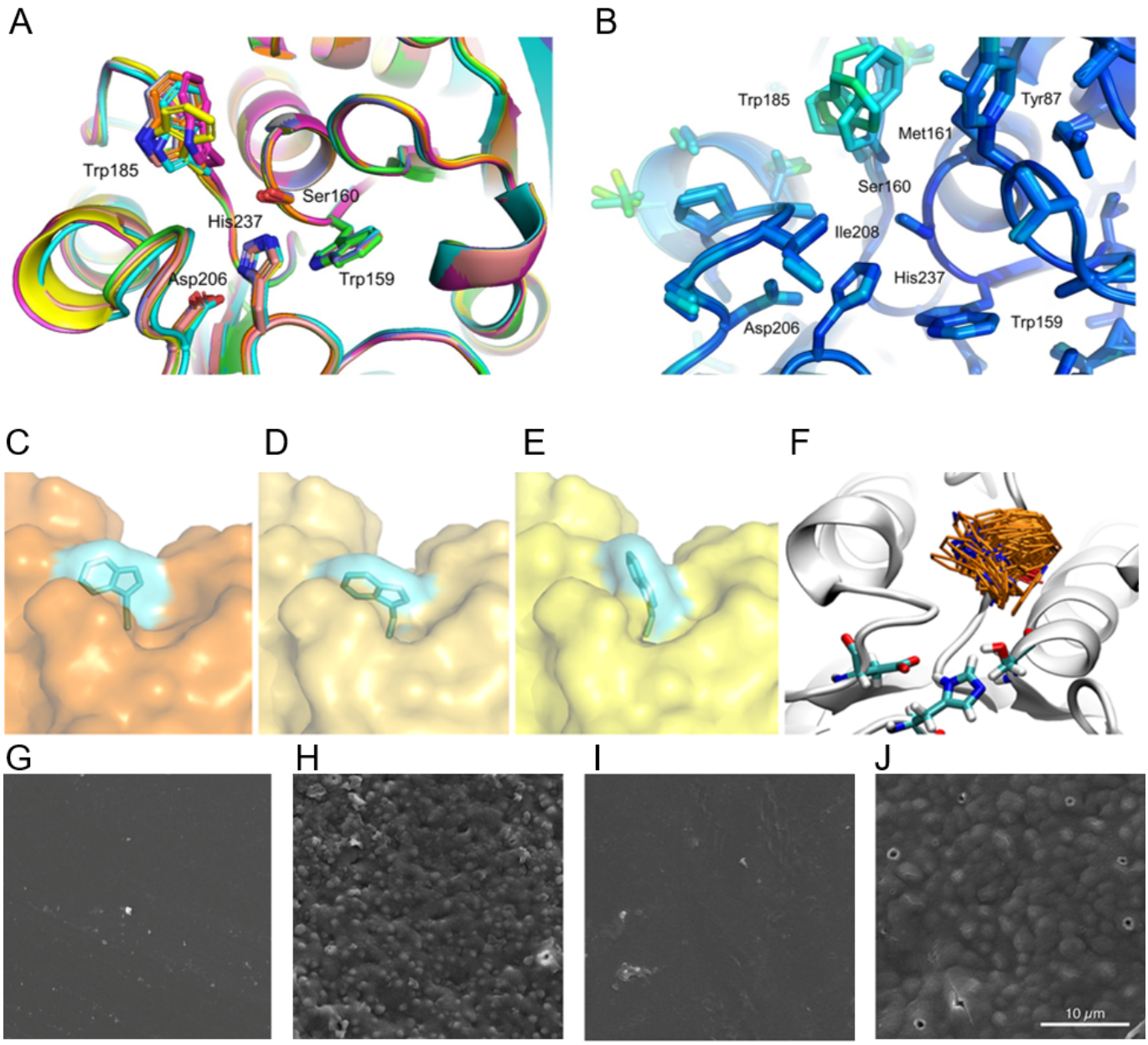



**Fig. S5.** Structural and functional analysis of key residues in PETase. (A) Superposition of the PETase structures (PDB IDs: 6EQE, 6EQF, 6EQG, and 6EQH). The RMSD of chains is 0.28 Å +/- 0.02. Overall, all the chains adopt the same fold with only slight variations in the loops. The sidechains of the three residues in the catalytic triad, Ser160, Asp206, and His 237 in addition to Trp159 and Trp185 are shown. All side-chain adopt the same rotameric state with the exception of Trp185, where two residues adopted slightly different $\chi$-2 orientations, possibly reflecting a degree of mobility. (B) Superposition of the three PETase domains from 6EQG with residues and sidechains colored according to B-factors. All the residues in the catalytic triad and Trp159 have considerably lower B-factors compared to Trp185. The increased B-factors indicate a greater mobility of this sidechain with respect to all of the other residues in the hydrophobic cleft and the catalytic triad. (C-E) Surface views of the binding cleft are shown from three independently refined chains in the asymmetric unit of crystal form 4, in space group $P2_1$ (PDB ID: 6EQG). The position of the Trp185 residue is highlighted in blue in each case. (F) MD simulation demonstrates a wide range of movement for residue Trp185 (orange). The catalytic triad residues, Ser160, Asp206, and His 237, are highlighted in cyan. (G) PET degradation after incubation with buffer only or (H) the W185A mutant PETase. (I) PEF degradation after incubation with buffer only or (J) the W185A mutant PETase.

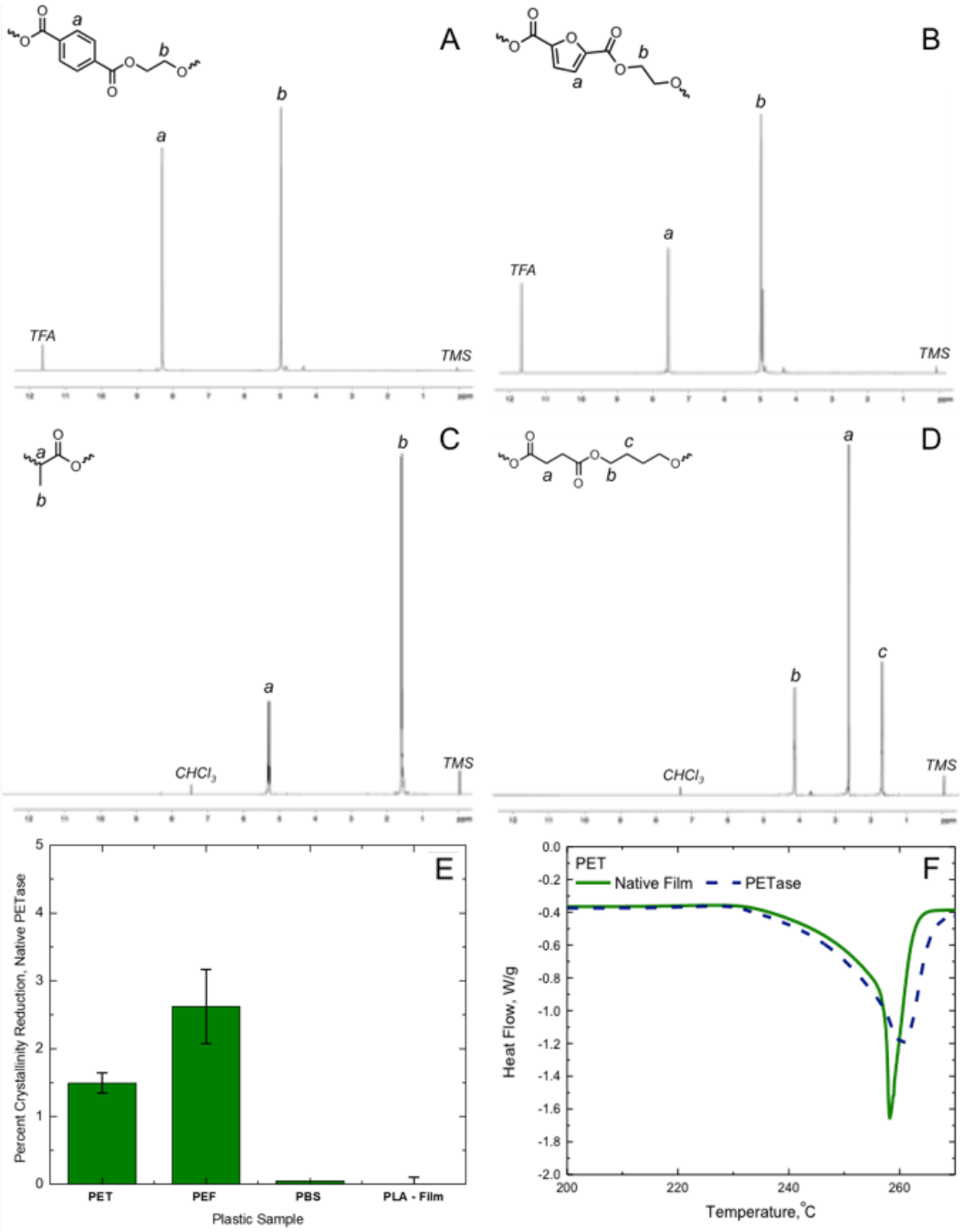



**Fig. S6.** Chemical analysis of polymer substrates. (A) $^{1}$H NMR spectra of the lab synthesized poly(ethylene terephthalate) (PET) used in the study. (B) $^{1}$H NMR spectra of the lab synthesized poly(ethylene 2,5 furandicarboxylate) (PEF) used in the study. (C) $^{1}$H NMR spectra of the lab synthesized poly(lactic acid) and (D) poly(butylene succinate. (E) Reduction in crystallinity in samples before and after digestion as determined by DSC. (F) Representative DSC trace for PET. After digestion, the melting transition is broadened indicating a reduction in both crystallinity and crystal domain size.

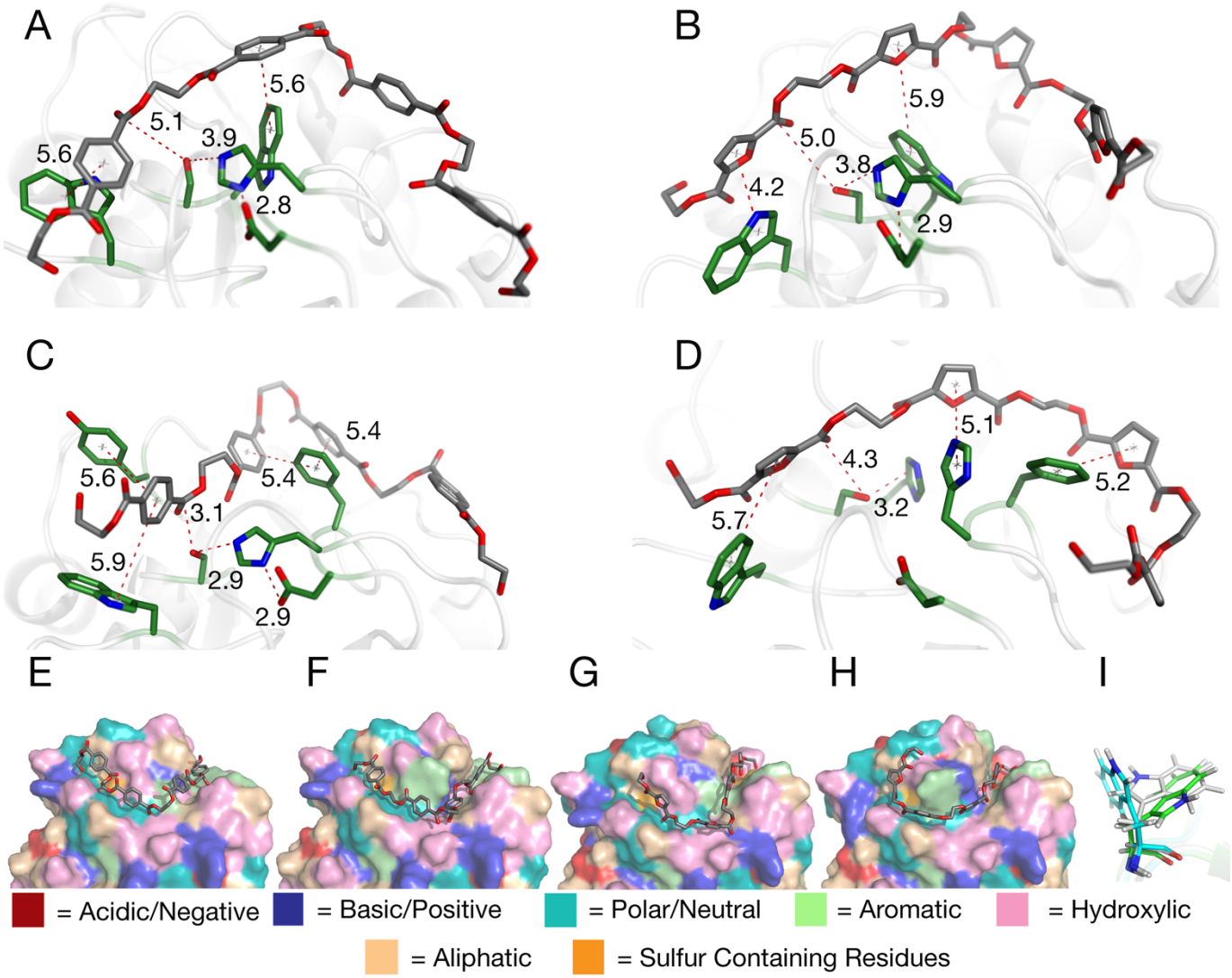



**Fig. S7.** Induced Fit Docking analysis. All distances given in ångströms. (A) Lowest energy, catalytically competent predicted pose of PET tetramer in wild-type PETase, XP score of -8.23 kcal/mol; catalytic triad is intact, W185 and W159 stabilize PET at optimal distances through parallel displaced and edge-to-face aromatic interactions, respectively. (B) Lowest energy, catalytically competent predicted pose of PEF tetramer in wild-type PETase, XP score of -9.07 kcal/mol; catalytic triad is intact, W185 and W159 stabilize PEF at optimal distances through parallel displaced and edge-to-face aromatic interactions, respectively. (C) Lowest energy, catalytically competent pose of PET tetramer in double mutant PETase, XP score of -11.25 kcal/mol; catalytic triad is intact. Here, PET is stabilized by four optimal aromatic contacts: edge-to-face to Trp185, parallel displaced to Tyr87, point-to-face to Phe238, parallel displaced to Phe238 (i.e., Phe238 participates in aromatic interactions from two terephthalate units). (D) Lowest energy, catalytically competent pose of PEF tetramer in double mutant PETase, XP score of -10.07 kcal/mol. Here, PEF is stabilized through three optimal aromatic interactions: parallel displaced to Trp185, parallel displaced toe Phe238, and a point-to-face interaction with His237. Due to the occupation of His237 with aromatic stabilization, it is not oriented within the catalytic triad, but rather His159 supports Ser160 via hydrogen bonding. This potential role of His159 as secondary hydrogen bonding support was independently observed by us through molecular dynamics simulations of the apo PETase double mutant. It should also be noted that His237 is within single bond rotation of catalytic triad participation, thus one could hypothesize it may assume this catalytic position after initial acyl-enzyme formation and protonation of His159. (E) Surface view of PET docked in the PETase wild-type structure, amino acids are colored according to their properties. (F) Surface view of PET docked in PETase double mutant structure. (G) Surface view of PEF docked in PETase wild type. (H) Surface view of PEF docked in PETase double mutant structure. E-H one can see PET and PEF occupy the same binding channel in both wild type and mutant proteins; also the S238F mutation changes the nature of the binding cleft. (I) Overlay of Trp185 position in PETase crystal structure (white), when PET is flexibly docked (green) and when PEF is flexibly docked (cyan). In response to PET/PEF binding, Trp185 rotates to provide optimal aromatic interactions for stabilization. In PETase crystal structure, the N-$C_\alpha$-$C_\beta$-C dihedral is -177.5°, whereas when PET is docked the same dihedral has a value of 98.4°, and -146.4° with PEF docked. This dihedral is also flexible in the double mutant structure, having values of 178.8° with PET and -155.5° with PEF flexibly docked. Trp185 is not the only flexible residue in the binding site, such rotations were also seen for H237, W159 (WT), H159 (mut), F238 (mut). The rotation of these residues illustrates the importance of modeling induced fit effects for predicting ligand binding modes. W185's flexibility was also captured with molecular dynamics simulations, **Fig. S5F**.

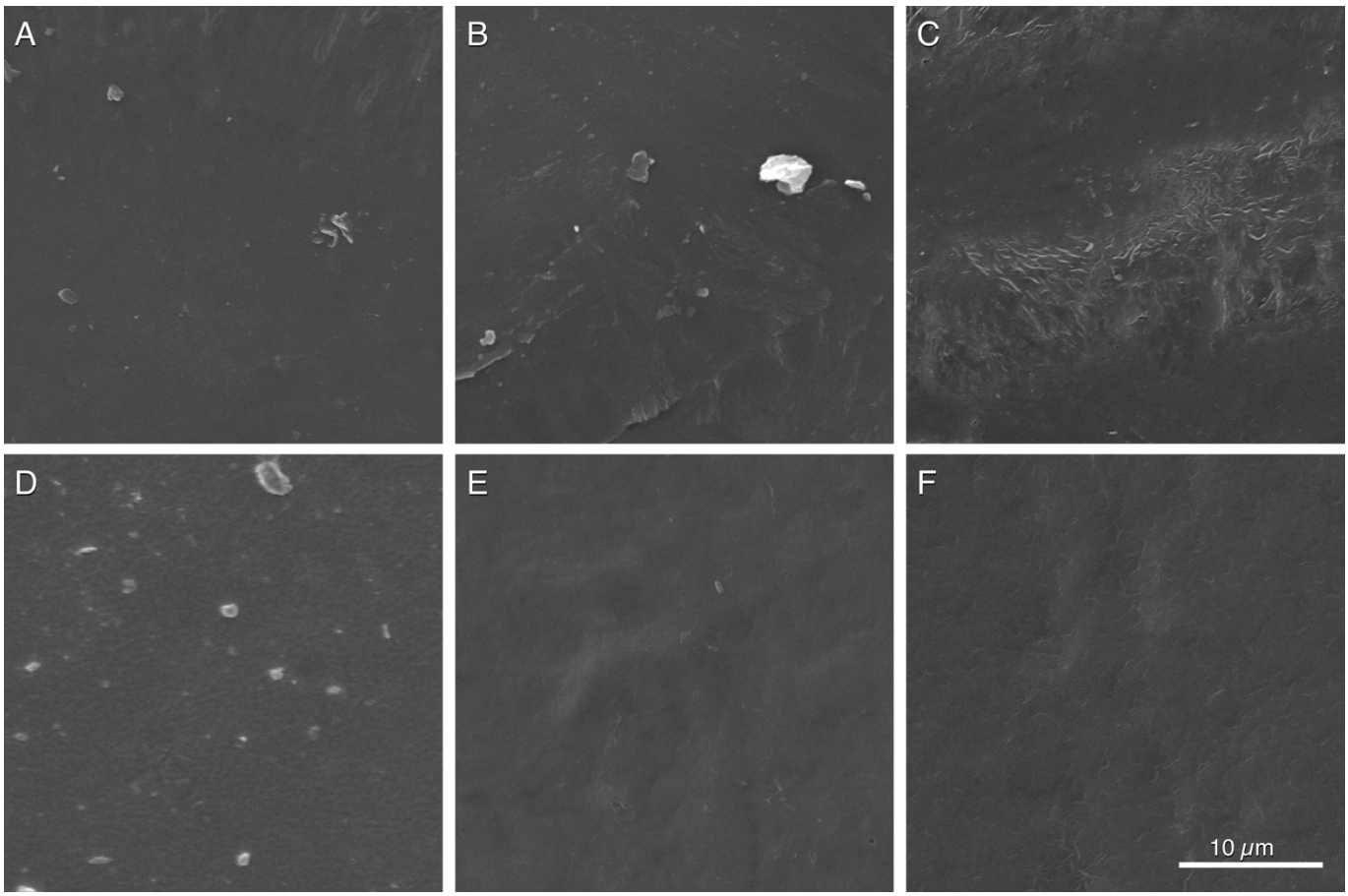



**Fig. S8.** Degradation analysis of PBS and PLA by PETase. (A) PBS coupon before incubation, (B) after incubation in buffer only, or (C) with PETase. (D) PLA film before incubation, (E) after incubation in buffer only, or (F) with PETase. All images are taken after 96 h of incubation at a PETase loading of 50 nM, pH 7.2 in phosphate buffer, or a buffer-only control. No surface pitting was observed.

## Table S1. Crystallographic data and refinement statistics including crystallization conditions

| | Long λ | Native 1 | Native 2 | Native 3 | Native 4 |
|---|---|---|---|---|---|
| **Data collection** | **(Form1)** | **(Form 2)** | **(Form 3)** | **(Form 4)** | **(Form 1)** |
| Symmetry | $C222_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_1$ | $C222_1$ |
| Wavelength | 2.4551 | 1.2398 | 0.9763 | 0.9763 | 0.9763 |
| Resolution Range (Å) | 82.3 - 1.70 | 84.54 - 0.92 | 68.24 – 1.70 | 46.5 - 1.8 | 49.8 – 1.58 |
| Unique reflections | 95250 | 154417 | 25470 | 68988 | 139895 |
| Completeness (%)[a] | 85.3 (74.1) | 99.9 (99.8) | 99.8 (98.3) | 97.7 (95.9) | 99.9 (99.9) |
| Anomalous Completeness (%)[a] | 81.3 (69.2) | | | | |
| *R*merge[b] | 0.052 (0.952) | 0.045 (0.510) | 0.107 (0.323) | 0.102 (0.537) | 0.049 (0.742) |
| CC(1/2)[c] | 0.999 (0.870) | 1.000 (0.852) | 0.991 (0.905) | 0.995 (0.776) | 0.999 (0.886) |
| Multiplicity[d] | 9.7 (8.5) | 6.5 (4.9) | 5.7 (5.7) | 3.4 (3.4) | 6.6 (6.5) |
| Anomalous Multiplicity[d] | 5.2 (4.5) | | | | |
| I/σI[a] | 18.8 (1.6) | 17.4 (2.2) | 9.8 (4.7) | 8.7 (2.3) | 18.2 (2.5) |
| | a = 52.56 Å, b = 234.31 Å, c = 164.53 Å | a = 51.11 Å, b = 51.39 Å, c = 84.54 Å | a = 51.35 Å, b = 64.13 Å, c = 68.24 Å | a = 51.11 Å, b = 51.39 Å, c = 84.54 Å, β = 93.19° | a = 52.70 Å, b = 234.13 Å, c = 165.12 Å |
| **Model Refinement** | | | | | |
| Resolution Range (Å) | 67.33 - 1.70 | 27.5 - 0.92 | 46.73 - 1.70 | 46.46 – 1.80 | 43.68 – 1.58 |
| No. of residues: | A: 261, B: 264, C: 263 | A: 265 | A: 269 | A: 262, B: 262, C: 263 | A: 263, B: 264, C: 262 |
| No. of water, ligands | 812, 2 Cl | 393, 2 Na, 2 Cl | 337, 1 Cl | 1153, 1 $SO_4$, 2 Cl | 990 |
| Rwork/Rfree (%)[e] | 18.76 (20.47) | 10.01 (11.01) | 14.92 (18.54) | 15.14 (18.38) | 16.91 (18.41) |
| B average[f] | 36.2 | 9.8 | 17.5 | 18.6 | 27.5 |
| Geometry bond, angles[g] | 0.002, 0.528 | 0.007, 1.104 | 0.005, 0.715 | 0.003, 0.598 | 0.007, 0.826 |
| Ramachandran[h] | 97.83, 0.0 | 98.1, 0.0 | 98.1, 0.0 | 97.95, 0.0 | 98.6, 0.0 |
| Molprobity Clash Score | 1.32 | 1.19 | 2.02 | 0.61 | 0.96 |
| Beamline | I23 | I23 | I03 | I03 | I03 |
| PDB ID[i] | 6EQD | 6EQE | 6EQF | 6EQG | 6EQH |

[a] Signal to noise ratio of intensities, highest resolution bin in brackets. [b] $R_m$ : $\sum h \sum i |I(h,i) - I(h)| / \sum h \sum i I(h,i)$ where $I(h,i)$ are symmetry-related intensities and $I(h)$ is the mean intensity of the reflection with unique index $h$. [c] $CC_{1/2}$ is the correlation coefficient of the mean intensities between two random half-datasets. [d] Multiplicity for unique reflections. [e] 5% of reflections were randomly selected for determination of the free R factor, prior to any refinement. [f] Temperature factors averaged for all atoms. [g] RMS deviations from ideal geometry for bond lengths and restraint angles (Engh and Huber). [h] Percentage of residues in the 'most favored region' of the Ramachandran plot and percentage of outliers (MOLPROBITY). [i] Protein Data Bank identifiers for coordinates.

Crystallography conditions: Long λ, 0.1M MIB (Malonate, Imidazole, Borate, pH 5.0), 25% PEG 1500. Native 1, 0.2M $MgCl_2$, 0.1M MES (pH 6.0), 20% PEG 6000. Native 2, 0.2M $NH_4Cl$, 0.1M MES (pH 6.0), 20% PEG 6000. Native 3, 0.2M Li $SO_4$, 0.1M Bis-Tris (pH 5.5), 25% PEG 3350. Native 4, 1.6M $NH_4SO_4$, 0.1M MES (pH 6.5).

**Table S2. Crystallinity determined by differential scanning calorimetry for plastics with PETase and mutations for the aromatic polyesters**

| Sample | Enzyme | Enthalpy before (J/g) | Enthalpy after (J/g) | Crystallinity before (%) | Crystallinity after (%) | Absolute crystallinity decrease (%) |
|---|---|---|---|---|---|---|
| PET[a] | PETase | 20.7 +/- 0.3 | 18.60 +/- 0.4 | 14.80 +/- 0.2 | 13.3 +/- 0.2 | 1.49 |
| | W159H,S238F | 20.7 +/- 0.3 | 12.70 +/- 0.4 | 14.80 +/- 0.2 | 9.18 +/- 0.2 | 5.62 |
| | W185A | 20.7 +/- 0.3 | 19.30 +/- 0.4 | 14.80 +/- 0.2 | 13.78 +/- 0.2 | 0.92 |
| PEF | PETase | 19.97 +/- 0.5 | 17.97 +/- 0.6 | 14.37 +/- 0.3 | 13.10 +/- 0.4 | 2.32 |
| | W159H,S238F | 19.72 +/- 0.5 | 16.31 +/- 0.6 | 14.37 +/- 0.3 | 11.65 +/- 0.4 | 3.75 |
| | W185A | 19.72 +/- 0.5 | 19.50 +/- 0.6 | 14.37 +/- 0.3 | 13.93+/- 0.4 | 1.47 |
| PBS[c] | PETase | 96.7 +/- 0.3 | 96.9 +/- 0.3 | 48.4[b] | 48.5[b] | 0 |
| PLA | PETase | 13.3 +/- 0.1 | 13.2 +/- 0.3 | 14.0 +/- 0.1 | 14.0 +/- 0.1 | 0 |

[a]On average cold crystallization for the PET samples is 24 J/g
[b]The error in PBS crystallinity is outside of significant figures
[c]Results for PETase double mutant similarly exhibit no change in measured properties

**Table S3: XP Descriptors for each of the discussed binding modes.** XP descriptors are an energetic decomposition of the Glide XP score, which itself is an estimate of the binding affinity. Thus, XP descriptors provide insight into which energetic terms contribute most significantly to binding free energy. All values are in kcal/mol. It can be seen $E_{vdW}$ contributes most significantly to all poses, and is even more favorable in mutant binding likely due to increased aromatic interactions with Phe238.

| Pose | XPScore | $E_{Hbond}$ | $E_{PhobEn}$(a) | $E_{PhobEnHBond}$(b) | $E_{RotPenal}$(c) | $E_{vdW}$ | $E_{Elec}$ | $E_{Sitemap}$(d) | $E_{Expospenal}$(e) |
|---|---|---|---|---|---|---|---|---|---|
| PET/WT | -8.23 | -3.56 | 0 | 0 | 0.30 | -5.88 | -1.05 | 0.68 | 2.68 |
| PET/Mut* | -11.25 | -3.00 | -0.35 | -1 | 0.30 | -6.61 | -0.93 | -0.73 | 1.08 |
| PEF/WT | -9.07 | -4.77 | 0 | 0 | 0.35 | -3.62 | -1.62 | -0.62 | 1.22 |
| PEF/Mut* | -10.07 | -2.70 | -0.35 | -1 | 0.35 | -4.70 | -1.20 | -0.90 | 0.43 |

*Mut here referring to S238F/W159H double mutant PETase structure.

(a) $E_{PhobEn}$ estimates energetic contribution from hydrophobic enclosure, or the hydrophobic effect

(b) $E_{PhobEnHBond}$ calculates contribution from a "buried" hydrogen bond, or hydrogen bond within a hydrophobic environment.

(c) $E_{RotPenal}$ is a value intrinsic to the ligand, it is a ligand penalty for the number of rotatable bonds.

(d) $E_{Sitemap}$ penalizes poses that align polar (non Hbonding)/hydrophobic or hydrophobic/hydrophilic regions (i.e., it penalizes alignment of non-complimentary regions between ligand and receptor). A decrease in $E_{Sitemap}$ indicates improved alignment between complimentary regions.

(e) $E_{Expospenal}$ penalty for solvent exposed ligand atoms; considering PETase's binding site is fairly solvent exposed this penalty cannot be avoided. Decrease in this value indicates a decrease in solvent exposure to the ligand.