nature portfolio

https://doi.org/10.1038/s41586-022-04599-z

Supplementary information

Machine learning-aided engineering of hydrolases for PET depolymerization

In the format provided by the authors and unedited

Supplementary Materials for:

Machine learning-aided engineering of hydrolases for PET depolymerization

Hongyuan Lu[1], Daniel J. Diaz[2], Natalie J. Czarnecki[1], Congzhi Zhu[1], Wantae Kim[1], Raghav Shroff[3,4], Daniel J. Acosta[3], Bradley R. Alexander[3], Hannah O. Cole[1,3], Yan Zhang[3], Nathaniel A. Lynd[1], Andrew D. Ellington[3], Hal S. Alper[1,*]

Affiliations

[1]McKetta Department of Chemical Engineering, The University of Texas at Austin, Austin, Texas 78712, United States

[2]Department of Chemistry, The University of Texas at Austin, Austin, Texas 78712, United States

[3]Department of Molecular Biosciences, The University of Texas at Austin, Austin, Texas 78712, United States

[4]DEVCOM ARL-South, Austin, Texas 78712, United States

*Corresponding Author: McKetta Department of Chemical Engineering, The University of Texas at Austin, 200 East Dean Keeton St., C0400, Austin, Texas 78712, halper@che.utexas.edu

Supplementary Materials Guide:

Supplementary Method: Sequences used in this study

Nucleotide sequence of the signal peptide—SPpstu (21 amino acids) from maltotetraose-forming amylase of *Pseudomonas stutzeri* MO-19[1]
atgagccacatcctgcgagccgccgtattggcggcgatgctgttgccgttgccgtccatggcc

Wild-type PETase

Nucleotide sequence of wild-type PETase without its original signal peptide
cagaccaatccgtatgcgcgcggtccgaatccgacagccgccagtttggaagcgagcgctggtccattcaccgttcgctcctttaccgtgagta gaccgagcggttatggcgctggcaccgtttactatccaacaaatgctgggggtaccgtgggcgccatagccatagttcccgggtatacggcac ggcagtcatcaattaaatggtggggaccgcgtctggcatcccacggtttcgtagtaattacaattgacacaaattccacgttagaccagccatcaa gtcggagttcgcaacaaatggccgcgctgcgccaggtggcgtcgttaaacggtacaagtagcagcccgatttacggaaaggtcgataccgct cgtatgggtgttatggggtggagtatgggaggtggaggctccctgatctctgctgctaacaacccttcgctgaaagcagcggcgcctcaagca ccatgggattcttcgacaaattttagttctgtaactgtgcccacgctgatcttcgcatgtgaaaacgatagtatagccccggtcaactcttcagcact tcctatctatgattctatgtcacgcaacgctaagcagtttctcgaaattaatggtggctcacattcctgtgcgaatagcggcaattctaaccaagcatt aatcggaaaaaaaggcgttgcatggatgaaacgttttatggacaatgatactaggtattctacttttgcctgcgagaacccgaatagcaccagagt gtctgattttcgtacagcgaattgcagcctcgagcaccaccaccaccaccac

Expressed amino acid sequence of wild-type PETase by *P. putida* KT2440
QTNPYARGPNPTAASLEASAGPFTVRSFTVSRPSGYGAGTVYYPTNAGGTVGAIAIVPGYT ARQSSIKWWGPRLASHGFVVITIDTNSTLDQPSSRSSQQMAALRQVASLNGTSSSPIYGKV DTARMGVMGWSMGGGGSLISAANNPSLKAAAPQAPWDSSTNFSSVTVPTLIFACENDSIA PVNSSALPIYDSMSRNAKQFLEINGGSHSCANSGNSNQALIGKKGVAWMKRFMDNDTRYS TFACENPNSTRVSDFRTANCSLEHHHHHH

ThermoPETase

Nucleotide sequence of ThermoPETase without its original signal peptide sequence
cagaccaatccgtatgcgcgcggtccgaatccgacagccgccagtttggaagcgagcgctggtccattcaccgttcgctcctttaccgtgagta gaccgagcggttatggcgctggcaccgtttactatccaacaaatgctgggggtaccgtgggcgccatagccatagttcccgggtatacggcac ggcagtcatcaattaaatggtggggaccgcgtctggcatcccacggtttcgtagtaattacaattgacacaaattccacgttagaccagccagaa agtcggagttcgcaacaaatggccgcgctgcgccaggtggcgtcgttaaacggtacaagtagcagcccgatttacggaaaggtcgataccgc tcgtatgggtgttatggggtggagtatgggaggtggaggctccctgatctctgctgctaacaacccttcgctgaaagcagcggcgcctcaagca ccatggcactcttcgacaaattttagttctgtaactgtgcccacgctgatcttcgcatgtgaaaacgatagtatagccccggtcaactcttcagcact tcctatctatgattctatgtcacgcaacgctaagcagtttctcgaaattaatggtggctcacattcctgtgcgaatagcggcaattctaaccaagcatt aatcggaaaaaaaggcgttgcatggatgaaacgttttatggacaatgatactaggtattctacttttgcctgcgagaacccgaatagcaccgcagt gtctgattttcgtacagcgaattgcagcctcgagcaccaccaccaccaccac

Expressed amino acid sequence of ThermoPETase by *P. putida* KT2440
QTNPYARGPNPTAASLEASAGPFTVRSFTVSRPSGYGAGTVYYPTNAGGTVGAIAIVPGYT ARQSSIKWWGPRLASHGFVVITIDTNSTLDQPESRSSQQMAALRQVASLNGTSSSPIYGKV DTARMGVMGWSMGGGGSLISAANNPSLKAAAPQAPWHSSTNFSSVTVPTLIFACENDSIA PVNSSALPIYDSMSRNAKQFLEINGGSHSCANSGNSNQALIGKKGVAWMKRFMDNDTRYS TFACENPNSTAVSDFRTANCSLEHHHHHH

DuraPETase

Nucleotide sequence of DuraPETase without its original signal peptide sequence
cagaccaatccgtatgcgcgcggtccgaatccgacagccgccagtttggaagcgagcgctggtccattcaccgttcgctcctttaccgtgagta gaccgagcggttatggcgctggcaccgtttactatccaacaaatgctgggggtaccgtgggcgccatagccatagttcccgggtatacggcac ggcagtcatcaattaaatggtggggaccgcgtctggcatcccacggtttcgtagtaattacaattgacacaaattccacgtttgactatccatcaag tcggagttcgcaacaaatggccgcgctgcgccaggtggcgtcgttaaacggtgacagtagcagcccgatttacggaaaggtcgataccgctc gtatgggtgttatggggcatagtatgggaggtggagcatccctgcgatctgctgctaacaacccttcgctgaaagcagcgattcctcaagcacc atgggattctcaaacaaattttagttctgtaactgtgcccacgctgatcttcgcatgtgaaaacgatagtatagccccggtcaactctcatgcacttc ctatctatgattctatgtcacgcaacgctaagcagtttctcgaaattaatggtggctcacattcctgtgcgaatagcggcaattctaaccaagcatta atcggaaaaaaaggcgttgcatggatgaaacgttttatggacaatgatactaggtattctacttttgcctgcgagaacccgaatagcaccgcagtg tctgattttcgtacagcgaattgcagcctcgagcaccaccaccaccaccac

Expressed amino acid sequence of DuraPETase by *P. putida* KT2440
QTNPYARGPNPTAASLEASAGPFTVRSFTVSRPSGYGAGTVYYPTNAGGTVGAIAIVPGYT ARQSSIKWWGPRLASHGFVVITIDTNSTFDYPSSRSSQQMAALRQVASLNGDSSSPIYGKV DTARMGVMGHSMGGGASLRSAANNPSLKAAIPQAPWDSQTNFSSVTVPTLIFACENDSIA PVNSHALPIYDSMSRNAKQFLEINGGSHSCANSGNSNQALIGKKGVAWMKRFMDNDTRY STFACENPNSTAVSDFRTANCSLEHHHHHH

Cut190

Nucleotide sequence of Cut190 without its original signal peptide sequence
caggacaatccatatgaacgtggacctgatcctaccgaggacagcatcgaggctatccgcggcccattcagtgtggctacagagcgcgttagc tccttcgcttctggtttcggcggcggtaccatctattatccccgtgaaaccgacgagggaacttttggagctgtcgccgtggcacctggatttacg gcgagccagggctctatgtcttggtacggggaacgcgttgcgtcgcaaggtttcatcgttttcacaattgacaccaacacccgtcttgaccaacc cggccaacgcggtcgtcaattattagctgctctggactacttggtcgagcgctcagatcgtaaggtacgtgaacgtttagacccaaatcgtctgg ctgtgatgggccacagtatgggcggaggcggttcattggaggcaacagtcatgcgtccctcgctgaaggcgtcgatccctttgaccccatgga atctggacaaaacttggggacaagtgcaggttcccacgtttattattggcgcggagttagacaccatcgcatccgtgcgtactcatgcgaaaccg ttctatgagtcacttccctcctccctgcccaaagcatatatggaacttgacggtgcgactcattttgcgccaaacattccgaataccacaatcgcta agtatgtaattagttggttaaaacgcttcgtggacgaggatactcgctactcccaattcctgtgtccaaaccctaccgatcgtgcaattgaggaata tcgtagtacatgtccttatctcgagcaccaccaccaccaccac

Expressed amino acid sequence of Cut190 by *E. coli*
QDNPYERGPDPTEDSIEAIRGPFSVATERVSSFASGFGGGTIYYPRETDEGTFGAVAVAPGF TASQGSMSWYGERVASQGFIVFTIDTNTRLDQPGQRGRQLLAALDYLVERSDRKVRERLD PNRLAVMGHSMGGGGSLEATVMRPSLKASIPLTPWNLDKTWGQVQVPTFIIGAELDTIAS VRTHAKPFYESLPSSLPKAYMELDGATHFAPNIPNTTIAKYVISWLKRFVDEDTRYSQFLCP NPTDRAIEEYRSTCPYLEHHHHHH

LCC

Nucleotide sequence of LCC without its original signal peptide sequence
agcaacccgtaccagcgtggcccgaatccgacccgcagcgcactgaccgcagatggcccgtttagcgtggcaacctacaccgtctcacgcct gtcagtctcgggttttggcggtggcgtgatttattacccgaccggcacgtctctgacgttcggtggcatcgcgatgagtccgggttataccgcag atgctagctctctggcatggctgggtcgtcgcctggcttcccatggctttgtggttctggtgattaacacgaattcacgtttcgattatccggacagc cgcgcctctcagctgagtgccgccctgaactacctgcgtaccagttccccgagcgccgttcgcgcacgtctggatgcaaatcgtctggcggttg ccggtcattctatgggtggcggtggcaccctgcgtattgcagaacaaaacccgagcctgaaagcggctgtcccgctgaccccgtggcacacc gataaaacgtttaataccagtgtcccggtgctgattgttggcgcagaagctgacaccgtggcgccggtttcgcagcatgccatcccgttttatcaa aacctgccgagcaccacgccgaaagtttacgtcgaactggataacgcatcgcacttcgctccgaatagcaacaatgcggccatttccgtttatac gatctcatggatgaaactgtgggtcgataatgacacccgttaccgccagttcctgtgtaatgtgaacgacccggctctgtccgacttccgcacca ataatcgccactgccaactcgagcaccaccaccaccaccac

Expressed amino acid sequence of LCC by *E. coli*
SNPYQRGPNPTRSALTADGPFSVATYTVSRLSVSGFGGGVIYYPTGTSLTFGGIAMSPGYTA DASSLAWLGRRLASHGFVVLVINTNSRFDYPDSRASQLSAALNYLRTSSPSAVRARLDAN RLAVAGHSMGGGGTLRIAEQNPSLKAAVPLTPWHTDKTFNTSVPVLIVGAEADTVAPVSQ HAIPFYQNLPSTTPKVYVELDNASHFAPNSNNAAISVYTISWMKLWVDNDTRYRQFLCNV NDPALSDFRTNNRHCQLEHHHHHH

ICCM

Nucleotide sequence of ICCM without its original signal peptide sequence
atgagcaacccgtaccagcgtggcccgaatccgacccgcagcgcactgaccgcagatggcccgtttagcgtggcaacctacaccgtctcac gcctgtcagtctcgggttttggcggtggcgtgatttattacccgaccggcacgtctctgacgttcggtggcatcgcgatgagtccgggttataccg cagatgctagctctctggcatggctgggtcgtcgcctggcttcccatggctttgtggttctggtgattaacacgaattcacgtttcgattatccggac agccgcgcctctcagctgagtgccgccctgaactacctgcgtaccagttccccgagcgccgttcgcgcacgtctggatgcaaatcgtctggcg gttgccggtcattctatgggtggcggtggcaccctgcgtattgcagaacaaaacccgagcctgaaagcggctgtcccgctgaccccgtggcac accgataaaacgtttaataccagtgtcccggtgctgattgttggcgcagaagctgacaccgtggcgccggtttcgcagcatgccatcccgttttat caaaacctgccgagcaccacgccgaaagtttacgtcgaactgtgcaacgcatcgcacattgctccgatgagcaacaatgcggccatttccgttt atacgatctcatggatgaaactgtgggtcgataatgacacccgttaccgccagttcctgtgtaatgtgaacgacccggctctgtgcgacttccgca ccaataatcgccactgccaactcgagcaccaccaccaccaccac

Expressed amino acid sequence of ICCM by *E. coli*
MSNPYQRGPNPTRSALTADGPFSVATYTVSRLSVSGFGGGVIYYPTGTSLTFGGIAMSPGY TADASSLAWLGRRLASHGFVVLVINTNSRFDYPDSRASQLSAALNYLRTSSPSAVRARLDA NRLAVAGHSMGGGGTLRIAEQNPSLKAAVPLTPWHTDKTFNTSVPVLIVGAEADTVAPVS QHAIPFYQNLPSTTPKVYVELCNASHIAPMSNNAAISVYTISWMKLWVDNDTRYRQFLCN VNDPALCDFRTNNRHCQLEHHHHHH

Supplementary Discussion

Supplementary Discussion 1:

Comparison between MutCompute and other deep learning models

While other unsupervised deep learning models (such as DeepSequence[2] and SeqDesign[3]) have also demonstrated the ability to guide protein engineering efforts, previous efforts have largely been sequence-based models that are either trained or fine-tuned on homologous sequences. Since MutCompute is explicitly trained on local structural and chemical features, it can accelerate the identification and experimental validation of amino acids never tried by nature, especially since the input data is a sequence-balanced dataset that intentionally removes homologous sequences

Specifically, MutCompute utilizes a type of unsupervised learning—self-supervision—that enables it to leverage the evolutionary signal observed in the functional proteins deposited in the PDB to guide learning. However, MutCompute is not the only unsupervised deep learning model in the literature that leverages evolution to guide protein engineering. Other unsupervised deep learning models such as DeepSequence[2] and SeqDesign[3], have also demonstrated the ability to guide protein engineering efforts and strongly correlate with deep mutational scanning experimental datasets.

However, DeepSequence[2] and SeqDesign[3] are trained with multiple sequence alignments (MSAs) and protein sequences as input, respectively, while MutCompute is trained on a voxelized representation of local protein chemistry. This modal difference in training data has several implications in how the data is sourced and pre-processed and the type of models trained.

First, these sequence-based models are trained on evolutionary data for a protein family; thus, for every target protein a new model must be trained on its homologous sequences. MutCompute training set, however, was designed to be agnostic to the target protein's family since it utilizes a 50% sequence similarity cut-off to intentionally remove homologous proteins.

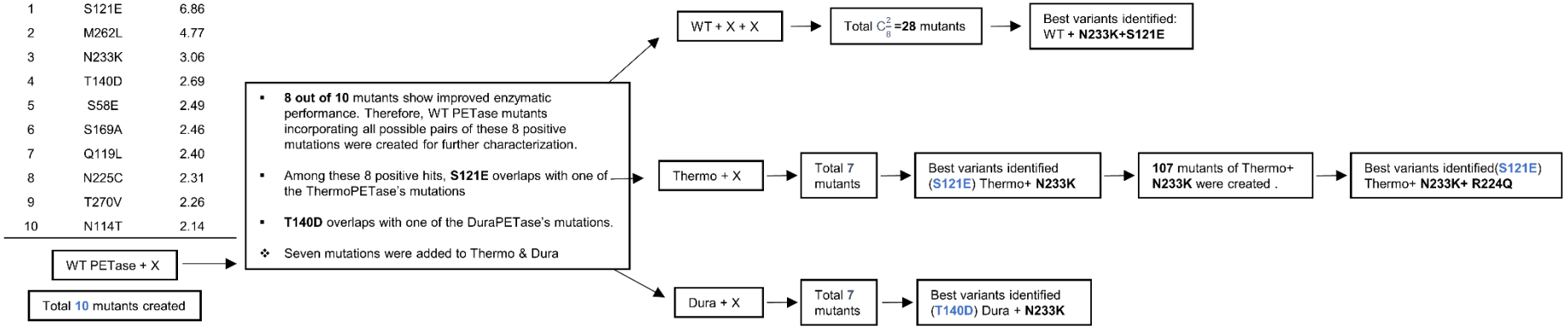
Second, a protein structure is much richer in information compared to a protein sequence. Thus, end-to-end training of language models, such as DeepSequence and SeqDesign, requires orders of magnitude more proteins since these models must first implicitly learn structural and chemical constraints from homologous sequences in order to inference functional changes accurately. However, these structural and chemical constraints are explicit in a protein structure, which not only enables us to train with just 19K protein structures but accelerates MutCompute's ability to learn protein chemistry and flag anomalous protein chemistry for mutagenesis experiments. So far, we have demonstrated the generalizability of the protein chemistry MutCompute learns by engineering proteins from a variety of protein families: BFP[4], TEM1[4], Bst polymerase[5], and now a PETase. Nonetheless, there is valuable information in homologous proteins, and we have yet to explore softening the sequence similarity cut-

off and/or transfer learning MutCompute to homologous structures of a target protein. Such modifications to our training procedure may improve the efficacy of MutCompute's designs translating from in silico to in carbo.

Third, sequence-based models cannot be easily tailored to guide the engineering of specificity at protein-ligand, antibody-antigen, and protein-nucleotide interactions because these other chemical entities are not a part of the protein sequence. This limitation is apparent in the SeqDesign paper[3] where they fit an autoregressive model to 1.2 M nanobody sequences in order to generate synthetic nanobody libraries enriched for stability and expression rather than to generate nanobody libraries enriched for antigen specificity, which is the real goal. MutCompute, however, can easily be adapted to provide mutagenesis designs at such interfaces because the input is voxelized protein chemistry and the training data is readily available via co-crystals in the PDB and can be supplemented with AlphaFold[6] and computational docking[7–9].

Supplementary Discussion 2: <u>Selecting mutations based on experimental catalytic activity measurements.</u>

Ranking	Prediction	avg_log_ratio
1	S121E	6.86
2	M262L	4.77
3	N233K	3.06
4	T140D	2.69
5	S58E	2.49
6	S169A	2.46
7	Q119L	2.40
8	N225C	2.31
9	T270V	2.26
10	N114T	2.14



A scheme for selecting mutations based on experimental evidence is provided as above. Initially, we chose the top ten ranked predictions based on crystal structure PDB: 5XJH (Extended Data Fig. 1a) and introduced them respectively into the wild-type PETase scaffold to generate ten single mutants. Experimental characterization of these variants showed that eight out of these ten predicted mutations confer improved thermostability and activity to the wild-type PETase scaffold. Notably, of such eight beneficial mutations, S121E and T140D are each overlapped with one of the mutations of ThermoPETase and DuraPETase. Subsequently, we paired two of the eight beneficial mutations to create all 28 possible double mutants of wild-type PETase. Meanwhile, the unique beneficial mutations were respectively introduced into ThermoPETase and DuraPETase scaffolds to generate 14 variants that contain two predictions from MutCompute. Among these 42 variants, the best variants identified for each scaffold all contain the predicted mutation-N233K. As the variant exhibiting the highest enzymatic activity, ThermoPETase^{N233K} was chosen as the template for further mutagenesis. Finally, a total of 107 variants of ThermoPETase^{N233K} were created by incorporating single or multiple mutations from the 14 top ranked predictions based on both crystal structures PDB: 5XJH and 6IJ6 as well as lower ranked predictions selected by a rational design strategy. After comparative analysis of the enzymatic performance of these 107 variants, ThermoPETase$^{N233K/R224Q}$ was identified as the best variant as it showed improved activity versus ThermoPETase^{N233K}. Given the synergetic interactions among the four predicted mutations that resulted in the best WT PETase, ThermoPETase and DuraPETase variants, S121E, N233K, R224Q and T140D were further selected for combinatorial assembly and follow-up analysis.

Supplementary Table

Probability distribution across all 20 amino acids																						
pdb_id	pos	wtAA	wt_prob	ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
6ij6	233	ASN	0.09%	1.E-05	1.E-02	9.E-04	2.E-04	2.E-05	5.E-03	4.E-03	3.E-07	6.E-04	1.E-03	5.E-04	1.E+00	4.E-04	2.E-04	1.E-06	8.E-05	3.E-04	2.E-05	1.E-04	9.E-04
6ij6	91	GLN	0.09%	9.E-06	6.E-05	6.E-05	9.E-07	3.E-05	9.E-04	3.E-05	6.E-08	2.E-07	1.E+00	4.E-05	2.E-04	1.E-04	3.E-09	2.E-07	2.E-05	6.E-03	1.E-08	1.E-08	1.E-02
6ij6	225	ASN	0.32%	3.E-04	9.E-06	3.E-03	7.E-07	1.E+00	2.E-05	9.E-07	4.E-07	4.E-07	1.E-05	1.E-04	3.E-07	1.E-04	3.E-08	2.E-07	2.E-04	2.E-02	4.E-09	2.E-09	1.E-03
5xjh	140	THR	0.41%	7.E-05	2.E-06	1.E-02	1.E+00	2.E-03	3.E-05	6.E-04	5.E-07	2.E-06	5.E-07	1.E-05	1.E-05	3.E-06	1.E-07	1.E-07	1.E-02	4.E-03	9.E-08	2.E-07	2.E-05
6ij6	61	SER	0.54%	4.E-04	9.E-04	3.E-02	2.E-01	2.E-03	1.E-02	9.E-02	9.E-06	9.E-03	2.E-02	2.E-02	1.E-03	1.E-03	1.E-03	2.E-05	5.E-03	3.E-01	2.E-04	2.E-03	3.E-01
5xjh	121	SER	0.70%	2.E-03	2.E-02	3.E-02	4.E-02	2.E-03	1.E-01	5.E-01	3.E-05	7.E-03	5.E-04	2.E-03	2.E-01	2.E-03	2.E-04	1.E-03	7.E-03	6.E-04	2.E-05	1.E-04	7.E-04
5xjh	58	SER	0.81%	4.E-03	2.E-02	3.E-02	1.E-01	1.E-03	6.E-02	5.E-01	9.E-05	3.E-03	7.E-03	8.E-02	1.E-01	2.E-03	5.E-04	4.E-05	8.E-03	4.E-03	1.E-04	3.E-04	7.E-03
5xjh	95	LYS	1.06%	2.E-04	8.E-04	1.E-02	2.E-02	2.E-03	2.E-01	6.E-01	2.E-07	2.E-05	4.E-02	1.E-03	1.E-02	2.E-03	9.E-07	5.E-07	9.E-03	4.E-03	4.E-07	9.E-07	1.E-01
5xjh	279	THR	1.98%	9.E-03	2.E-02	1.E-01	5.E-01	2.E-03	3.E-02	2.E-01	2.E-04	9.E-03	5.E-03	3.E-03	3.E-02	2.E-03	1.E-03	1.E-04	8.E-02	2.E-02	2.E-04	8.E-04	9.E-03
5xjh	263	ASP	2.25%	7.E-05	2.E-05	5.E-02	2.E-02	4.E-04	5.E-02	8.E-03	9.E-07	9.E-01	7.E-07	2.E-05	5.E-04	6.E-03	2.E-04	2.E-07	9.E-05	1.E-04	2.E-06	5.E-04	6.E-07
6ij6	58	SER	2.30%	8.E-01	1.E-02	6.E-03	3.E-02	4.E-04	8.E-03	1.E-01	1.E-03	4.E-04	2.E-04	6.E-03	2.E-02	6.E-04	1.E-04	5.E-05	2.E-02	4.E-03	4.E-05	2.E-04	3.E-03
6ij6	53	ARG	2.82%	1.E-04	3.E-02	1.E-02	4.E-04	7.E-05	3.E-01	7.E-02	1.E-06	3.E-05	3.E-04	7.E-03	6.E-01	8.E-03	5.E-06	1.E-06	2.E-04	1.E-04	5.E-06	6.E-06	2.E-04
6ij6	34	ARG	2.99%	4.E-07	3.E-02	2.E-02	2.E-04	8.E-05	3.E-03	5.E-04	3.E-07	1.E-01	3.E-02	8.E-01	2.E-02	4.E-03	6.E-04	6.E-07	2.E-05	1.E-03	4.E-05	9.E-05	5.E-04
6ij6	208	ILE	3.02%	4.E-04	4.E-03	4.E-04	3.E-04	3.E-04	6.E-04	8.E-04	3.E-06	7.E-05	3.E-02	5.E-04	1.E-03	5.E-04	3.E-05	1.E-05	6.E-04	7.E-02	1.E-05	1.E-05	9.E-01
6ij6	59	ARG	3.25%	2.E-02	3.E-02	6.E-02	5.E-02	6.E-03	6.E-02	1.E-01	3.E-04	3.E-03	4.E-03	1.E-03	9.E-02	3.E-03	4.E-04	2.E-04	2.E-01	3.E-01	5.E-05	5.E-04	6.E-02
6ij6	224	ARG	3.52%	1.E-05	4.E-02	1.E-02	3.E-04	1.E-05	4.E-01	1.E-02	4.E-07	3.E-03	6.E-06	2.E-01	4.E-01	1.E-02	1.E-03	5.E-07	8.E-06	3.E-06	2.E-03	1.E-03	5.E-07
5xjh	262	MET	3.74%	5.E-06	8.E-06	1.E-02	9.E-05	2.E-03	3.E-04	9.E-05	5.E-07	3.E-05	2.E-01	7.E-01	5.E-05	4.E-02	3.E-06	8.E-07	7.E-06	6.E-04	3.E-07	5.E-07	5.E-02
6ij6	136	SER	4.15%	2.E-02	3.E-05	7.E-03	1.E-01	2.E-02	5.E-03	4.E-02	7.E-05	2.E-03	1.E-04	6.E-04	2.E-04	5.E-04	4.E-04	1.E-05	4.E-02	7.E-01	2.E-05	2.E-04	9.E-02

5xjh	233	ASN	4.51%	6.E-04	3.E-03	5.E-02	7.E-03	5.E-04	2.E-02	1.E-02	1.E-06	7.E-02	2.E-05	3.E-04	8.E-01	4.E-04	1.E-03	7.E-07	2.E-03	1.E-03	1.E-05	2.E-03	6.E-05
6ij6	186	HIS	4.75%	1.E-05	2.E-06	2.E-01	8.E-01	3.E-03	2.E-04	1.E-03	4.E-08	5.E-02	7.E-07	1.E-05	2.E-05	4.E-05	5.E-04	2.E-07	7.E-04	3.E-05	1.E-04	3.E-04	9.E-06
5xjh	225	ASN	4.83%	2.E-03	2.E-04	5.E-02	3.E-06	6.E-01	2.E-04	5.E-06	1.E-06	9.E-06	8.E-06	4.E-04	3.E-06	3.E-04	7.E-08	1.E-06	4.E-02	3.E-01	2.E-08	2.E-08	2.E-04
5xjh	59	ARG	5.74%	7.E-03	6.E-02	2.E-01	6.E-02	3.E-03	1.E-01	2.E-01	2.E-04	2.E-02	3.E-03	3.E-03	9.E-02	6.E-03	3.E-03	1.E-04	5.E-02	2.E-01	2.E-04	3.E-03	3.E-02
5xjh	212	ASN	6.08%	3.E-01	1.E-02	6.E-02	3.E-01	2.E-03	6.E-03	2.E-02	7.E-04	4.E-04	2.E-04	3.E-03	3.E-02	9.E-04	1.E-04	3.E-05	3.E-01	7.E-03	2.E-05	1.E-04	9.E-04
6ij6	37	ASN	6.92%	4.E-05	5.E-05	7.E-02	9.E-01	1.E-04	1.E-05	2.E-04	5.E-07	9.E-05	5.E-06	3.E-06	2.E-04	5.E-06	2.E-06	8.E-07	5.E-04	6.E-05	2.E-07	2.E-06	2.E-05
6ij6	33	MET	7.37%	3.E-05	4.E-03	7.E-03	4.E-03	4.E-04	4.E-01	4.E-01	4.E-06	4.E-02	6.E-03	7.E-02	6.E-02	7.E-02	4.E-04	4.E-05	6.E-04	3.E-03	2.E-05	5.E-04	1.E-03
6ij6	279	THR	7.58%	7.E-03	1.E-02	4.E-02	2.E-01	2.E-03	2.E-02	3.E-01	6.E-05	1.E-02	1.E-02	3.E-03	8.E-02	3.E-03	2.E-03	9.E-05	1.E-01	8.E-02	2.E-04	1.E-03	3.E-02
5xjh	119	GLN	7.67%	1.E-05	5.E-03	1.E-03	3.E-04	1.E-04	8.E-02	7.E-03	2.E-04	7.E-03	3.E-05	1.E-01	1.E-01	5.E-01	1.E-01	1.E-06	2.E-05	5.E-05	3.E-03	3.E-02	8.E-06
5xjh	125	SER	8.20%	4.E-02	4.E-03	6.E-03	3.E-02	2.E-02	2.E-02	4.E-02	5.E-05	2.E-04	7.E-03	2.E-03	1.E-02	2.E-03	7.E-05	5.E-05	8.E-02	2.E-01	3.E-05	8.E-05	5.E-01
6ij6	212	ASN	8.26%	2.E-01	3.E-02	8.E-02	1.E-01	4.E-03	3.E-02	1.E-01	3.E-04	2.E-02	9.E-03	3.E-03	7.E-02	2.E-03	3.E-03	5.E-05	2.E-01	5.E-03	2.E-04	2.E-03	1.E-02
6ij6	87	TYR	8.43%	8.E-08	5.E-01	7.E-06	1.E-06	2.E-06	1.E-03	1.E-04	4.E-08	1.E-02	1.E-05	3.E-03	4.E-01	4.E-03	2.E-02	5.E-08	2.E-07	1.E-07	2.E-03	8.E-02	2.E-07
5xjh	82	ALA	8.52%	9.E-02	1.E-07	1.E-05	2.E-06	5.E-03	4.E-07	7.E-07	5.E-04	9.E-08	2.E-08	2.E-08	7.E-08	3.E-07	2.E-08	2.E-06	9.E-01	2.E-04	1.E-09	1.E-08	2.E-06
5xjh	117	LEU	8.68%	7.E-05	6.E-03	4.E-01	7.E-03	3.E-02	1.E-01	2.E-01	2.E-05	2.E-02	5.E-04	9.E-02	1.E-01	3.E-02	8.E-04	1.E-05	2.E-03	3.E-02	2.E-05	2.E-04	3.E-04
6ij6	104	HIS	9.20%	7.E-09	1.E-06	2.E-05	1.E-06	1.E-07	1.E-05	7.E-06	3.E-10	9.E-02	4.E-08	5.E-08	7.E-07	4.E-06	6.E-02	6.E-10	4.E-08	2.E-08	5.E-04	9.E-01	3.E-08
5xjh	213	SER	9.23%	2.E-02	2.E-02	3.E-02	1.E-01	1.E-03	3.E-02	2.E-01	9.E-05	5.E-04	5.E-03	2.E-02	7.E-02	3.E-03	2.E-04	5.E-05	9.E-02	3.E-01	8.E-05	4.E-04	3.E-02
6ij6	240	ALA	9.40%	9.E-02	2.E-05	1.E-03	9.E-05	8.E-01	3.E-04	1.E-04	1.E-05	1.E-05	3.E-04	7.E-05	4.E-05	2.E-04	1.E-06	3.E-05	7.E-03	6.E-02	8.E-07	9.E-07	2.E-02
6ij6	46	SER	9.48%	8.E-01	7.E-03	3.E-03	3.E-02	1.E-03	7.E-03	6.E-02	4.E-03	1.E-04	5.E-05	5.E-04	3.E-02	5.E-04	3.E-05	9.E-05	9.E-02	3.E-03	1.E-05	3.E-05	3.E-04
6ij6	270	THR	10.14%	2.E-05	2.E-03	2.E-03	6.E-04	3.E-04	5.E-03	4.E-03	3.E-07	7.E-05	2.E-01	1.E-03	1.E-02	8.E-04	4.E-06	2.E-06	1.E-04	1.E-01	1.E-06	3.E-06	6.E-01
6ij6	183	ALA	12.26%	1.E-01	3.E-04	2.E-01	1.E-02	3.E-01	2.E-02	3.E-02	8.E-05	1.E-04	8.E-05	2.E-05	9.E-05	2.E-02	5.E-06	1.E-04	2.E-01	1.E-01	4.E-06	8.E-06	1.E-02
6ij6	110	THR	12.30%	7.E-05	1.E-06	9.E-03	1.E-04	9.E-04	5.E-05	3.E-05	6.E-07	3.E-06	7.E-01	7.E-04	5.E-05	2.E-05	3.E-08	9.E-07	2.E-05	1.E-01	2.E-08	8.E-08	2.E-01
5xjh	270	THR	12.72%	5.E-03	1.E-04	3.E-03	6.E-04	7.E-03	4.E-03	3.E-04	3.E-05	3.E-05	3.E-02	1.E-03	5.E-04	6.E-04	8.E-06	6.E-03	5.E-03	1.E-01	8.E-07	6.E-06	8.E-01
6ij6	238	SER	13.39%	8.E-01	1.E-07	2.E-05	8.E-05	8.E-05	1.E-07	1.E-07	9.E-02	8.E-08	4.E-10	6.E-09	2.E-07	3.E-08	7.E-09	1.E-07	1.E-01	8.E-06	1.E-09	4.E-09	4.E-09
5xjh	124	SER	13.51%	9.E-01	2.E-09	2.E-07	1.E-07	6.E-05	4.E-08	2.E-08	7.E-06	5.E-08	7.E-12	3.E-10	2.E-09	2.E-08	1.E-09	3.E-09	1.E-01	9.E-07	3.E-11	8.E-10	5.E-10
6ij6	95	LYS	13.62%	5.E-07	9.E-04	1.E-03	1.E-04	3.E-05	2.E-01	6.E-01	2.E-09	9.E-04	3.E-05	6.E-02	1.E-01	3.E-03	7.E-06	1.E-08	5.E-06	9.E-06	3.E-07	3.E-06	4.E-06
5xjh	274	GLU	14.23%	2.E-06	1.E-03	3.E-03	3.E-04	9.E-04	5.E-02	1.E-01	8.E-08	1.E-02	3.E-03	7.E-02	3.E-02	6.E-01	8.E-02	6.E-07	8.E-06	7.E-04	4.E-05	6.E-03	8.E-03

6ij6	119	GLN	14.67%	8.E-07	2.E-02	7.E-05	3.E-05	1.E-05	1.E-01	2.E-03	5.E-06	5.E-03	2.E-05	1.E-02	1.E-01	2.E-01	3.E-01	2.E-07	9.E-07	2.E-06	1.E-02	2.E-01	1.E-06
5xjh	51	THR	15.48%	3.E-02	8.E-03	4.E-02	5.E-02	3.E-03	6.E-02	5.E-01	3.E-04	3.E-04	7.E-05	5.E-03	9.E-02	3.E-03	4.E-05	1.E-05	2.E-02	2.E-01	8.E-06	3.E-05	2.E-03
5xjh	41	ALA	16.64%	2.E-01	1.E-02	1.E-02	4.E-02	9.E-04	3.E-02	1.E-01	5.E-01	5.E-04	6.E-05	4.E-03	4.E-02	2.E-03	3.E-04	5.E-04	5.E-02	6.E-03	7.E-05	2.E-04	2.E-04
5xjh	92	SER	17.08%	8.E-01	2.E-03	9.E-04	1.E-02	2.E-04	2.E-03	2.E-02	4.E-05	7.E-06	1.E-05	3.E-05	2.E-03	3.E-05	2.E-06	5.E-05	2.E-01	1.E-03	1.E-06	4.E-06	1.E-04
5xjh	63	TYR	17.23%	4.E-09	7.E-06	5.E-07	7.E-09	2.E-07	2.E-06	8.E-08	1.E-09	5.E-03	3.E-09	3.E-07	9.E-07	7.E-05	8.E-01	2.E-10	1.E-08	1.E-08	2.E-04	2.E-01	8.E-09
5xjh	133	GLN	17.50%	4.E-06	3.E-05	4.E-02	2.E-02	1.E-04	2.E-01	7.E-01	1.E-07	8.E-04	2.E-03	6.E-03	1.E-03	1.E-03	2.E-05	8.E-07	1.E-05	1.E-02	2.E-06	1.E-05	8.E-03
6ij6	236	SER	18.79%	1.E-01	3.E-04	7.E-02	6.E-01	1.E-03	6.E-04	3.E-04	4.E-04	1.E-05	7.E-06	1.E-05	4.E-04	3.E-05	7.E-07	1.E-05	2.E-01	5.E-04	1.E-06	9.E-07	3.E-05
5xjh	292	GLU	19.20%	4.E-03	3.E-02	1.E-01	7.E-02	4.E-03	8.E-02	2.E-01	9.E-04	6.E-02	2.E-02	1.E-01	2.E-01	2.E-02	2.E-02	3.E-04	3.E-02	3.E-02	1.E-03	5.E-03	1.E-02
5xjh	208	ILE	19.53%	6.E-04	9.E-03	3.E-03	2.E-03	2.E-03	1.E-02	2.E-02	4.E-05	3.E-04	2.E-01	1.E-03	1.E-02	5.E-03	2.E-04	5.E-05	2.E-03	2.E-01	1.E-04	1.E-04	6.E-01
6ij6	269	SER	20.31%	5.E-01	3.E-04	6.E-02	1.E-01	8.E-03	1.E-02	7.E-03	3.E-05	2.E-03	1.E-03	2.E-04	2.E-03	2.E-04	2.E-05	7.E-06	2.E-01	4.E-02	4.E-06	3.E-05	8.E-03
6ij6	190	ASN	21.04%	5.E-07	8.E-06	2.E-01	8.E-01	3.E-05	4.E-05	1.E-04	3.E-08	4.E-07	3.E-07	5.E-07	2.E-05	1.E-06	2.E-08	2.E-08	4.E-05	1.E-05	1.E-08	1.E-08	2.E-06
6ij6	41	ALA	22.07%	2.E-01	4.E-02	3.E-02	6.E-02	6.E-04	9.E-02	4.E-01	1.E-04	9.E-03	6.E-04	6.E-03	1.E-01	2.E-03	1.E-03	6.E-05	3.E-02	3.E-02	5.E-04	1.E-03	2.E-03
5xjh	154	MET	22.55%	5.E-07	5.E-03	4.E-04	1.E-05	1.E-05	7.E-01	5.E-02	5.E-08	1.E-04	5.E-06	1.E-03	3.E-02	2.E-01	4.E-04	3.E-08	2.E-07	7.E-07	2.E-04	2.E-04	2.E-07
5xjh	207	SER	22.65%	8.E-02	1.E-02	2.E-01	4.E-01	8.E-03	8.E-03	4.E-02	5.E-04	5.E-03	2.E-04	5.E-04	6.E-03	2.E-03	6.E-04	7.E-05	2.E-01	1.E-02	8.E-05	7.E-04	1.E-03
5xjh	73	ASN	22.79%	5.E-06	4.E-06	2.E-01	8.E-01	2.E-04	2.E-04	2.E-03	2.E-07	3.E-05	3.E-06	3.E-06	4.E-05	3.E-06	2.E-07	4.E-07	4.E-04	2.E-05	3.E-08	2.E-07	1.E-05
6ij6	207	SER	23.23%	2.E-01	1.E-02	1.E-01	4.E-01	4.E-03	5.E-03	1.E-02	2.E-03	7.E-03	4.E-04	5.E-04	8.E-03	8.E-04	2.E-03	9.E-05	2.E-01	7.E-03	9.E-04	2.E-03	2.E-03
6ij6	261	PHE	23.26%	1.E-08	4.E-06	1.E-03	4.E-07	6.E-07	9.E-05	4.E-05	3.E-09	4.E-01	8.E-08	1.E-07	8.E-06	1.E-04	2.E-01	2.E-09	2.E-08	2.E-08	3.E-01	1.E-01	3.E-09
5xjh	190	ASN	23.32%	2.E-05	4.E-03	2.E-01	8.E-01	2.E-04	1.E-03	2.E-03	4.E-06	6.E-04	7.E-06	1.E-04	2.E-03	2.E-04	9.E-05	1.E-06	6.E-04	1.E-04	1.E-04	9.E-05	4.E-05
6ij6	165	GLY	25.56%	7.E-01	1.E-07	2.E-07	3.E-07	1.E-05	7.E-08	5.E-08	3.E-01	2.E-08	3.E-09	2.E-08	8.E-08	3.E-08	1.E-08	8.E-06	4.E-03	1.E-06	1.E-09	3.E-09	4.E-08
6ij6	125	SER	26.36%	9.E-02	3.E-03	1.E-03	4.E-03	6.E-03	1.E-02	2.E-02	1.E-04	3.E-04	2.E-03	2.E-03	1.E-02	1.E-03	2.E-04	6.E-05	3.E-01	4.E-01	5.E-05	2.E-04	2.E-01
6ij6	175	SER	26.87%	2.E-01	4.E-03	4.E-02	4.E-01	1.E-03	3.E-03	4.E-02	4.E-04	4.E-04	1.E-05	8.E-04	9.E-03	5.E-04	8.E-05	5.E-04	3.E-01	1.E-02	1.E-05	8.E-05	8.E-05
5xjh	37	ASN	27.55%	7.E-07	2.E-04	3.E-01	7.E-01	9.E-05	7.E-05	1.E-03	7.E-08	1.E-03	7.E-04	2.E-05	1.E-03	4.E-05	2.E-05	4.E-07	4.E-05	7.E-04	7.E-07	1.E-05	9.E-04
5xjh	287	ALA	27.60%	3.E-01	4.E-03	9.E-03	2.E-01	7.E-03	4.E-03	2.E-02	2.E-03	4.E-03	2.E-05	2.E-04	2.E-02	1.E-03	3.E-04	3.E-05	5.E-01	6.E-03	1.E-04	3.E-04	8.E-04
6ij6	63	TYR	27.81%	8.E-09	5.E-06	3.E-06	2.E-08	6.E-07	1.E-05	2.E-07	3.E-09	2.E-02	8.E-09	1.E-06	1.E-06	2.E-04	7.E-01	6.E-10	3.E-08	5.E-08	3.E-04	3.E-01	2.E-08
6ij6	231	GLU	28.31%	3.E-06	3.E-04	3.E-03	3.E-05	5.E-03	7.E-01	3.E-01	4.E-08	2.E-06	2.E-06	1.E-03	5.E-03	8.E-03	1.E-08	1.E-08	9.E-05	1.E-06	6.E-08	9.E-09	5.E-07
6ij6	133	GLN	28.50%	3.E-06	2.E-04	4.E-02	1.E-01	9.E-05	3.E-01	5.E-01	3.E-07	5.E-03	6.E-04	5.E-03	4.E-03	2.E-03	7.E-05	1.E-06	1.E-05	1.E-03	8.E-06	7.E-05	6.E-03

5xjh	159	TRP	28.75%	1.E-10	3.E-05	3.E-07	1.E-07	9.E-09	8.E-06	4.E-07	5.E-11	3.E-02	3.E-10	9.E-07	3.E-05	3.E-05	3.E-01	5.E-11	6.E-10	3.E-10	3.E-01	3.E-01	4.E-10
6ij6	132	ARG	29.00%	3.E-05	3.E-01	4.E-03	1.E-03	8.E-05	4.E-01	3.E-02	7.E-07	6.E-02	7.E-05	5.E-04	8.E-02	1.E-02	3.E-02	1.E-06	4.E-05	4.E-04	6.E-02	2.E-02	2.E-04
5xjh	175	SER	29.89%	5.E-01	8.E-03	1.E-02	1.E-01	2.E-03	1.E-02	4.E-02	1.E-04	3.E-04	2.E-05	2.E-03	3.E-02	8.E-04	8.E-05	1.E-05	3.E-01	1.E-02	1.E-05	7.E-05	3.E-04

Supplementary Table 1 | Disfavored PETase residues flagged by MutCompute from the wild-type and ThermoPETase crystal structures. MutCompute outputs a probability distribution that describes the likelihood of each of the 20 canonical amino acids to be the wild-type amino acid for the surrounding chemical environment. A disfavored residue is defined as a residue where the amino acid with the highest predicted probability is not the wild-type amino acid. Here, a 30% wild-type probability cutoff was used to down select disfavored residues.

Ranking	position	wtAA	prAA	wt_prob	pred_prob	avg_log_ratio
1	121	SER	GLU	0.11%	61.20%	6.86
2	262	MET	LEU	4.12%	66.51%	4.77
3	233	ASN	LYS	4.86%	55.93%	3.06
4	140	THR	ASP	14.23%	75.20%	2.69
5	58	SER	GLU	5.22%	45.81%	2.49
6	169	SER	ALA	10.29%	89.60%	2.46
7	119	GLN	LEU	6.06%	54.65%	2.40
8	225	ASN	CYS	7.94%	78.08%	2.31
9	270	THR	VAL	8.65%	72.13%	2.26
10	114	ASN	THR	9.53%	76.00%	2.14
11	91	GLN	ILE	10.03%	53.89%	2.14
12	207	SER	ASP	15.83%	37.06%	1.78
13	212	ASN	ALA	6.17%	33.75%	1.73
14	59	ARG	ASN	6.26%	34.64%	1.69
15	136	SER	LEU	5.99%	28.02%	1.66
16	279	THR	GLU	6.85%	26.04%	1.66
17	168	ILE	LEU	28.37%	71.44%	1.65
18	263	ASP	ASN	19.55%	32.91%	1.65
19	154	MET	GLN	10.19%	46.02%	1.64
20	190	ASN	ASP	21.64%	77.84%	1.62
21	201	PHE	LEU	24.70%	67.93%	1.61
22	124	SER	ALA	25.46%	74.53%	1.32
23	208	ILE	VAL	18.96%	49.48%	1.23
24	117	LEU	LYS	9.84%	25.37%	1.21
25	95	LYS	GLU	20.94%	44.44%	1.11
26	73	ASN	ASP	28.55%	70.65%	1.06
27	53	ARG	LYS	29.40%	66.87%	1.00
28	274	GLU	MET	22.29%	46.47%	0.99
29	67	THR	VAL	29.74%	50.17%	0.97
30	125	SER	ALA	18.70%	38.47%	0.59
31	213	SER	THR	27.53%	40.39%	0.35
32	146	TYR	HIS	27.67%	44.27%	0.34
33	88	THR	THR	22.54%	22.54%	0.00
34	172	ASN	ASN	25.92%	25.92%	0.00
35	187	SER	SER	27.74%	27.74%	0.00
36	292	GLU	GLU	29.43%	29.43%	0.00
37	152	ALA	GLY	22.79%	35.05%	-0.10

Supplementary Table 2a | Predictions (based on wild-type PETase) ranked by fold change in the probabilities between the predicted and the wild-type amino acid. Fold change predictions are provided as a means of down-selecting potential mutations.

Ranking	position	wtAA	prAA	wt_prob	pred_prob	avg_log_ratio
1	91	GLN	ILE	0.09%	98.29%	7.42
2	233	ASN	LYS	0.09%	97.29%	7.39
3	225	ASN	CYS	0.32%	98.01%	5.86
4	61	SER	THR	0.54%	31.85%	3.78
5	208	ILE	VAL	3.02%	88.68%	3.57
6	58	SER	ALA	2.30%	78.28%	3.54
7	34	ARG	LEU	2.99%	78.89%	3.41
8	240	ALA	CYS	9.40%	81.59%	3.24
9	186	HIS	ASP	4.75%	77.21%	3.20
10	224	ARG	GLN	3.52%	39.61%	3.14
11	53	ARG	LYS	2.82%	56.07%	2.99
12	270	THR	VAL	10.14%	63.14%	2.91
13	136	SER	THR	4.15%	66.37%	2.82
14	37	ASN	ASP	6.92%	92.96%	2.65
15	104	HIS	TYR	9.20%	85.21%	2.33
16	46	SER	ALA	9.48%	75.76%	2.31
17	95	LYS	GLU	13.62%	57.80%	2.22
18	59	ARG	THR	3.25%	31.60%	2.18
19	87	TYR	ARG	8.43%	50.29%	2.14
20	238	SER	ALA	13.39%	78.03%	2.09
21	236	SER	ASP	18.79%	61.12%	1.82
22	110	THR	ILE	12.30%	69.39%	1.75
23	33	MET	GLN	7.37%	36.75%	1.69
24	165	GLY	ALA	25.56%	74.08%	1.55
25	279	THR	GLU	7.58%	31.98%	1.44
26	119	GLN	PHE	14.67%	29.22%	1.35
27	190	ASN	ASP	21.04%	78.93%	1.33
28	63	TYR	PHE	27.81%	69.99%	1.32
29	212	ASN	ALA	8.26%	23.24%	1.10
30	269	SER	ALA	20.31%	54.77%	0.98
31	231	GLU	GLN	28.31%	69.43%	0.93
32	133	GLN	GLU	28.50%	52.67%	0.80
33	183	ALA	CYS	12.26%	30.00%	0.74
34	175	SER	ASP	26.87%	42.64%	0.71
35	41	ALA	GLU	22.07%	38.13%	0.55
36	261	PHE	HIS	23.26%	37.42%	0.52
37	125	SER	THR	26.36%	39.71%	0.47
38	207	SER	ASP	23.23%	44.12%	0.46
39	132	ARG	GLN	29.00%	41.70%	0.37
40	277	ASN	ASN	23.99%	23.99%	0.00

Supplementary Table 2b | Predictions (based on ThermoPETase) ranked by fold change in the probabilities between the predicted and the wild-type amino acid. Fold change predictions are provided as a means of down-selecting potential mutations.

Sample number	Postconsumer Plastic products	Initial mass (mg)	Crystallinity %	Time for complete degradation (days)	Category	Mn kg/mol	Mw kg/mol	Đ	Degradation rate (mM/h)
#1	Allergy relief tablets packaging	8.12 ± 0.28	1.18% ± 0.02%	2.1	Medication packaging	30.2	55.4	1.83	1.09
#2	Hanging hooks packaging	7.9 ± 0.08	1.21% ± 0.09%	2.1	Household goods packaging	31.6	56.2	1.78	1.11
#3	Coffee cap	10.88 ± 0.48	1.23% ± 0.20%	3.7	Beverage packaging	30.7	53.9	1.76	0.88
#4	Toothbrush packaging	11.85 ± 0.4	1.30% ± 0.14%	3.7	Household goods packaging	33.1	61.2	1.85	0.99
#5	Tangyuan Container	8.37 ± 0.25	1.40% ± 0.14%	2.9	Food packaging	29.2	50.9	1.74	0.86
#6	Strawberries container	12.62 ± 0.09	1.42% ± 0.29%	4.4	Food packaging	29.0	50.9	1.76	0.86
#7	Razor blade container	5.36 ± 0.06	1.44% ± 0.25%	1.8	Household goods packaging	29.8	57.9	1.94	0.84
#8	Baby spinach container	10.58 ± 0.23	1.50% ± 0.21%	3.2	Food packaging	33.2	59.9	1.80	1.00
#9	Staple remover packaging	7.47 ± 0.06	1.54% ± 0.13%	2.1	Office supplies packaging	32.1	56.8	1.77	1.05

Supplementary Table 3 | Mass, crystallinity %, molecular weights (Mn, Mw), polydispersity indices (Đ), time for complete degradation, and degradation rate of various pc-PET films by FAST-PETase. The circular pc-PET films (6 mm in diameter) were hole-punched from 51 different post-consumer plastic products used in the packaging of food, beverages, medications, office supplies, household goods and cosmetics available at local grocery store chains (Walmart, Costco, and HEB). The pc-PET films were hydrolysed by serial treatment with FAST-PETase at 50 ºC until the films were completed degraded. The enzyme solution (200 nM of FAST-PETase in 100mM KH_2PO_4-NaOH (pH 8.0) buffer) was replenished every 24 hours. The crystallinity % of the intact pc-PET films was determined by DSC. The initial mass of the films was determined gravimetrically by a digital scale. Both DSC and gravimetric measurements were conducted in triplicate. Means ± s.d. (n=3) are shown.

Supplementary Table 3 continued

Sample number	Postconsumer Plastic products	Initial mass (mg)	Crystallinity %	Time for complete degradation (days)	Category	Mn kg/mol	Mw kg/mol	Đ	Degradation rate (mM/h)
#10	Hole puncher packaging	8.44 ± 0.17	1.55% ± 0.22%	2.8	Office supplies packaging	35.4	66.6	1.88	0.79
#11	Soy sauce container	8.57 ± 0.45	1.65% ± 0.08%	2.6	Food packaging	33.3	60.6	1.82	0.96
#12	Blueberries container	8.8 ± 0.08	1.65% ± 0.21%	2.8	Food packaging	27.2	50.3	1.85	1.01
#13	Soft-picks packaging	10.56 ± 0.35	1.68% ± 0.30%	2.9	Household goods packaging	32.6	59.8	1.83	1.12
#14	Lotion packaging	12.52 ± 0.52	1.68% ± 0.06%	4.2	Food packaging	27.1	48.2	1.78	0.91
#15	Tomato container	10.9 ± 0.18	1.73% ± 0.27%	3.7	Food packaging	28.2	52.3	1.85	0.91
#16	Egg container	12.81 ± 0.11	2.00% ± 0.16%	3.9	Food packaging	23.3	45.3	1.94	0.98
#17	Beancake container	6.69 ± 0.22	2.01% ± 0.03%	1	Food packaging	30.9	54.7	1.77	1.43
#18	Pumpkin pie container	14.16 ± 0.15	2.14% ± 0.21%	4.6	Food packaging	36.4	62.4	1.71	1.01
#19	Cookies container	9.58 ± 0.14	2.19% ± 0.22%	2.9	Food packaging	36.1	62.5	1.73	0.91
#20	Marinated chicken container	9.23 ± 0.03	2.19% ± 0.07%	2.8	Food packaging	26.0	52.4	2.02	0.98
#21	Blackberries container	11.71 ± 0.25	2.20% ± 0.17%	3.8	Food packaging	24.0	44.1	1.84	0.90

Supplementary Table 3 continued

Sample number	Postconsumer Plastic products	Initial mass (mg)	Crystallinity %	Time for complete degradation (days)	Category	Mn kg/mol	Mw kg/mol	Đ	Degradation rate (mM/h)
#22	Cleanser packaging	13.69 ± 0.66	2.28% ± 0.24%	5.2	Cosmetics packaging	36.2	66.3	1.83	0.73
#23	Strawberry almond pound cake	11.49 ± 0.2	2.29% ± 0.13%	3.9	Food packaging	24.8	61.5	2.48	0.92
#24	Instant noddle container	6.96 ± 0.06	2.45% ± 0.17%	2.3	Food packaging	28.1	50.9	1.81	0.84
#25	Eyedrop packaging	11.9 ± 0.31	2.49% ± 0.26%	3.8	Medication packaging	33.4	61.1	1.83	0.90
#26	Citrucel caplets outer-packaging	9.4 ± 0.27	2.53% ± 0.22%	3.4	Medication packaging	32.7	57.8	1.77	0.75
#27	Light bulb packaging	9.42 ± 0.82	2.53% ± 1.64%	2.6	Household goods packaging	32.7	59.9	1.83	0.97
#28	Salted egg pastry container	8.89 ± 0.31	2.56% ± 0.16%	2.3	Food packaging	28.7	53.7	1.87	1.08
#29	Ball pen packaging	12.87 ± 0.05	2.56% ± 0.24%	7	Office supplies packaging	33.7	62.3	1.85	0.66
#30	Watermelon container	8.82 ± 0.69	2.61% ± 1.05%	2.5	Food packaging	38.0	70.8	1.86	0.99
#31	Car odor-eliminating air freshener packaging	11.03 ± 0.35	2.67% ± 0.30%	3	Household goods packaging	31.3	55.1	1.76	1.09
#32	Croissant container	20.35 ± 0.25	2.71% ± 0.14%	6.1	Food packaging	32.5	58.3	1.79	0.98
#33	Razor blade container	14.95 ± 0.2	2.90% ± 0.16%	4.2	Household goods packaging	28.7	52.4	1.83	1.01

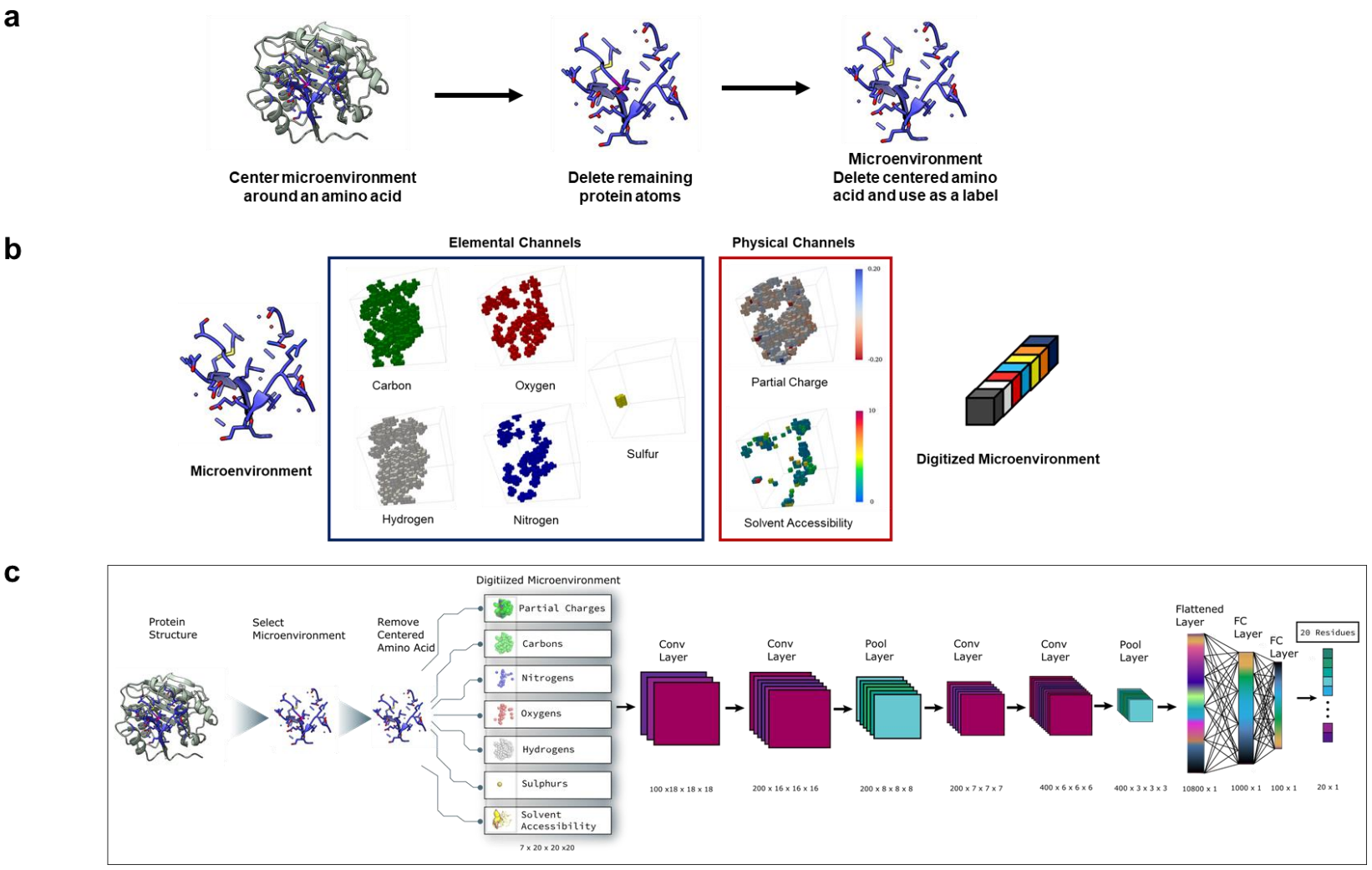
Supplementary Table 3 continued

Sample number	Postconsumer Plastic products	Initial mass (mg)	Crystallinity %	Time for complete degradation (days)	Category	Mn kg/mol	Mw kg/mol	Đ	Degradation rate (mM/h)
#34	Floss picks packaging	6.5 ± 0.44	2.93% ± 0.02%	1.9	Household goods packaging	36.8	67.7	1.84	0.90
#35	Seedweed container	7.44 ± 1.14	3.00% ± 0.61%	3	Food packaging	28.3	50.2	1.77	0.67
#36	Hand soap packaging	9.13 ± 0.09	3.06% ± 0.30%	2.7	Household goods packaging	29.8	52.7	1.77	0.97
#37	Boba milk cup	8.98 ± 0.54	3.21% ± 0.27%	2.6	Beverage packaging	33.2	57.5	1.73	0.98
#38	Muffin container	4.55 ± 0.4	3.42% ± 0.45%	3.8	Food packaging	30.9	56.1	1.82	0.48
#39	EZ-Cup disposable filters container	7.19 ± 0.2	3.47% ± 0.67%	2.5	Beverage packaging	35.7	64.1	1.80	0.86
#40	Cookies container	10.75 ± 0.06	3.56% ± 0.66%	4.1	Food packaging	39.7	68.7	1.73	0.79
#41	Medjool dates container	14.96 ± 0.04	3.57% ± 0.16%	4.6	Food packaging	31.8	57.0	1.79	0.89
#42	LED light bulb container	8.47 ± 0.24	3.58% ± 0.19%	2.2	Household goods packaging	31.2	58.0	1.86	1.09
#43	Color pens packaging	12.96 ± 0.15	3.72% ± 0.24%	6.7	Office supplies packaging	36.0	66.5	1.85	0.59
#44	Apple container	14.03 ± 0.62	4.09% ± 0.18%	4.7	Food packaging	31.4	56.2	1.79	0.85
#45	Tuekey breast container	17.11 ± 0.08	4.56% ± 0.26%	5.1	Food packaging	34.5	62.7	1.82	1.03

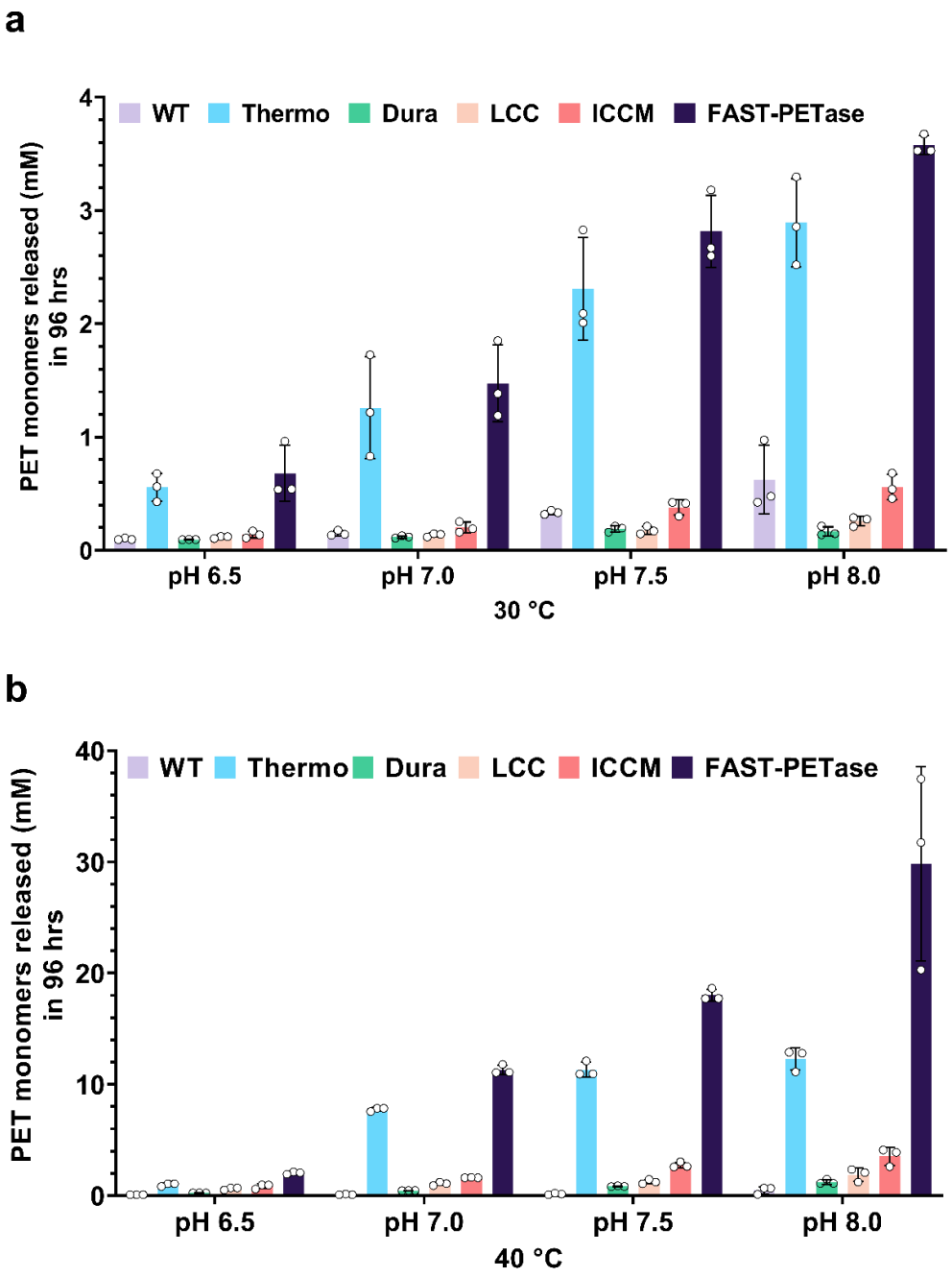
Supplementary Table 3 continued

Sample number	Postconsumer Plastic products	Initial mass (mg)	Crystallinity %	Time for complete degradation (days)	Category	Mn kg/mol	Mw kg/mol	Đ	Degradation rate (mM/h)
#46	Ginger container	9.5 ± 0.71	4.69% ± 1.08%	2.9	Food packaging	31.6	55.7	1.76	0.89
#47	Q-tip container	11.53 ± 0.1	4.85% ± 0.36%	3	Household goods packaging	35.1	64.1	1.83	1.11
#48	Kiwi fruit container	16.07 ± 0.26	4.85% ± 1.08%	4.8	Food packaging	33.1	57.9	1.75	1.00
#49	Sunscreen packaging	6.06 ± 0.22	5.30% ± 0.18%	2	Cosmetics packaging	28.8	50.9	1.77	0.89
#50	Shrimp container	10.94 ± 0.21	5.79% ± 0.11%	3.2	Food packaging	32.6	58.8	1.80	1.10
#51	Doll container	21.64 ± 0.36	6.24% ± 0.25%	6.8	Toy packaging	32.2	59.8	1.86	0.93

Supplementary Figures

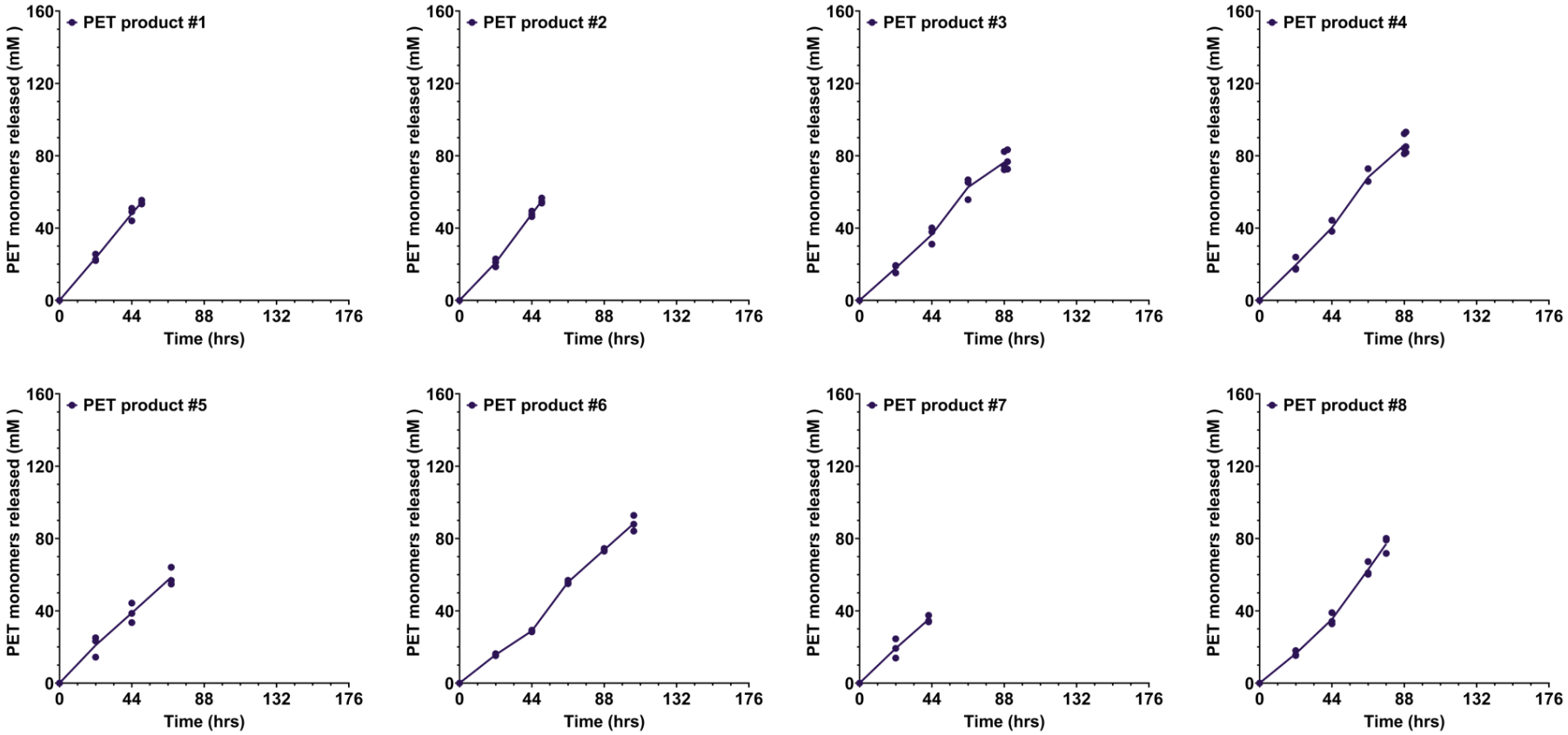


Supplementary Fig. 1 | Schematic diagram of MutCompute. a. Creating a microenvironment: MutCompute begins by centering itself on the alpha carbon of a particular residue in the protein and filters all peptide atoms within a 20 angstrom cube (the orientation of the cube is normalized with respect to the protein backbone). In the filtering process, we create an artificial, self-supervised label by excluding all atoms that belong the center residue. **b.** Encoding the microenvironment: The filtered atoms are then encoded into a 7-channel voxelated representation with a voxel resolution of $1A^3$. **c.** Running MutCompute on a Microenvironment: The 7-channel voxelated representation of a microenvironment is then passed to the CNN model, MutCompute. The model can be broken into 2 parts: Feature extraction and classification. The feature extraction portion consist of convolutional and max pooling layers and is then flattened into a 1D-vector before being passed to the classification layers of the model. The output is a probability mass function of the likelihood each of the 20 amino acids was the amino acid in the center of the microenvironment. We do this process for every residue in the protein to identify residues for mutagenesis.

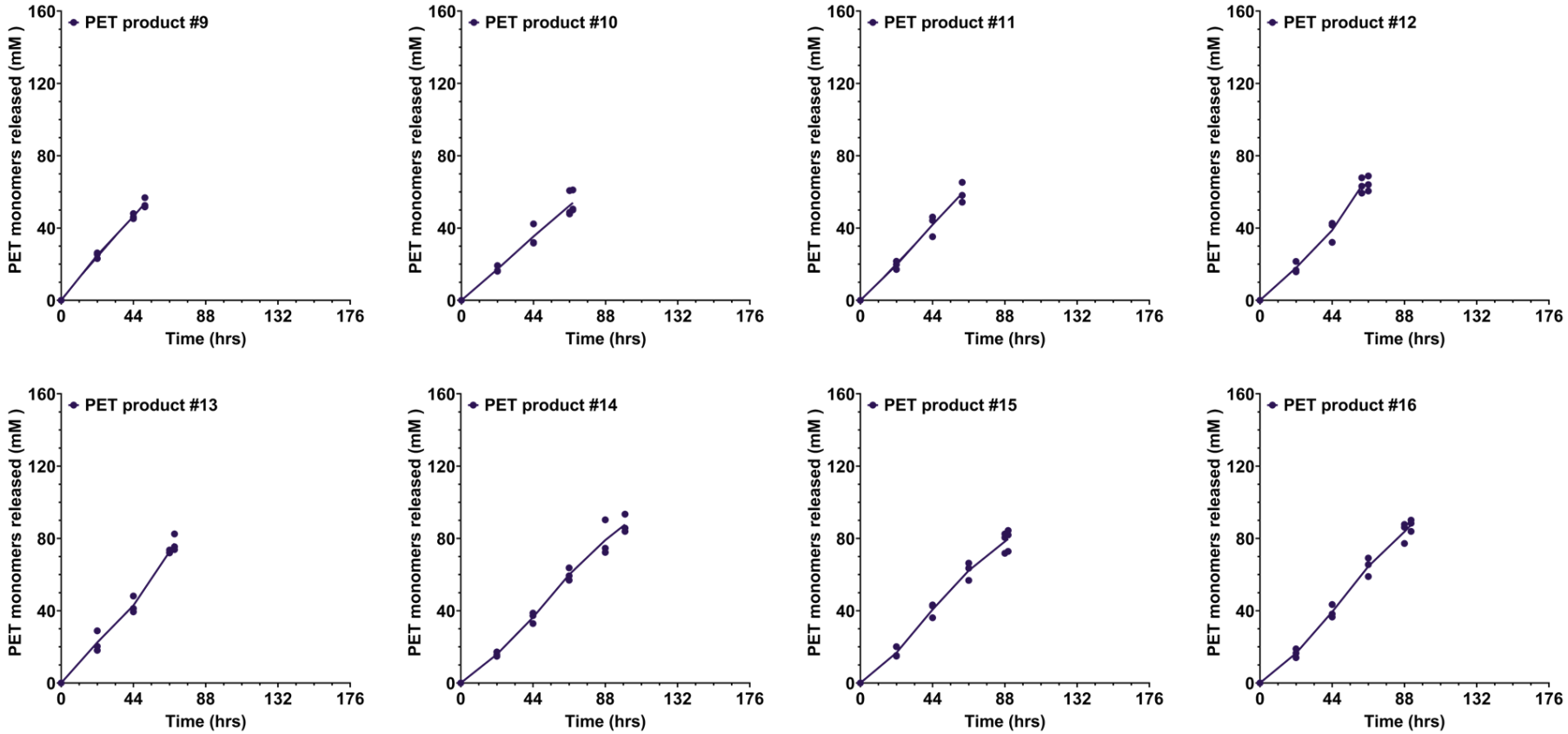


Supplementary Fig. 2 | The PET-hydrolytic activity of FAST-PETase outperformed various PHEs at mild temperatures and modest pH. Comparison of PET-hydrolytic activity of FAST-PETase, wild-type PETase (WT), ThermoPETase (Thermo), DuraPETase (Dura), LCC and ICCM **across** a range of pH (6.5 – 8.0) at reaction temperatures of 30 ºC. (**a.**) and 40 ºC (**b.**). PET-hydrolytic activity was evaluated by measuring the amount of PET monomers (the sum of TPA and MHET) released from hydrolyzing gf-PET film by the tested enzymes after 96 hrs of reaction time. All measurement were conducted in triplicate (n=3).

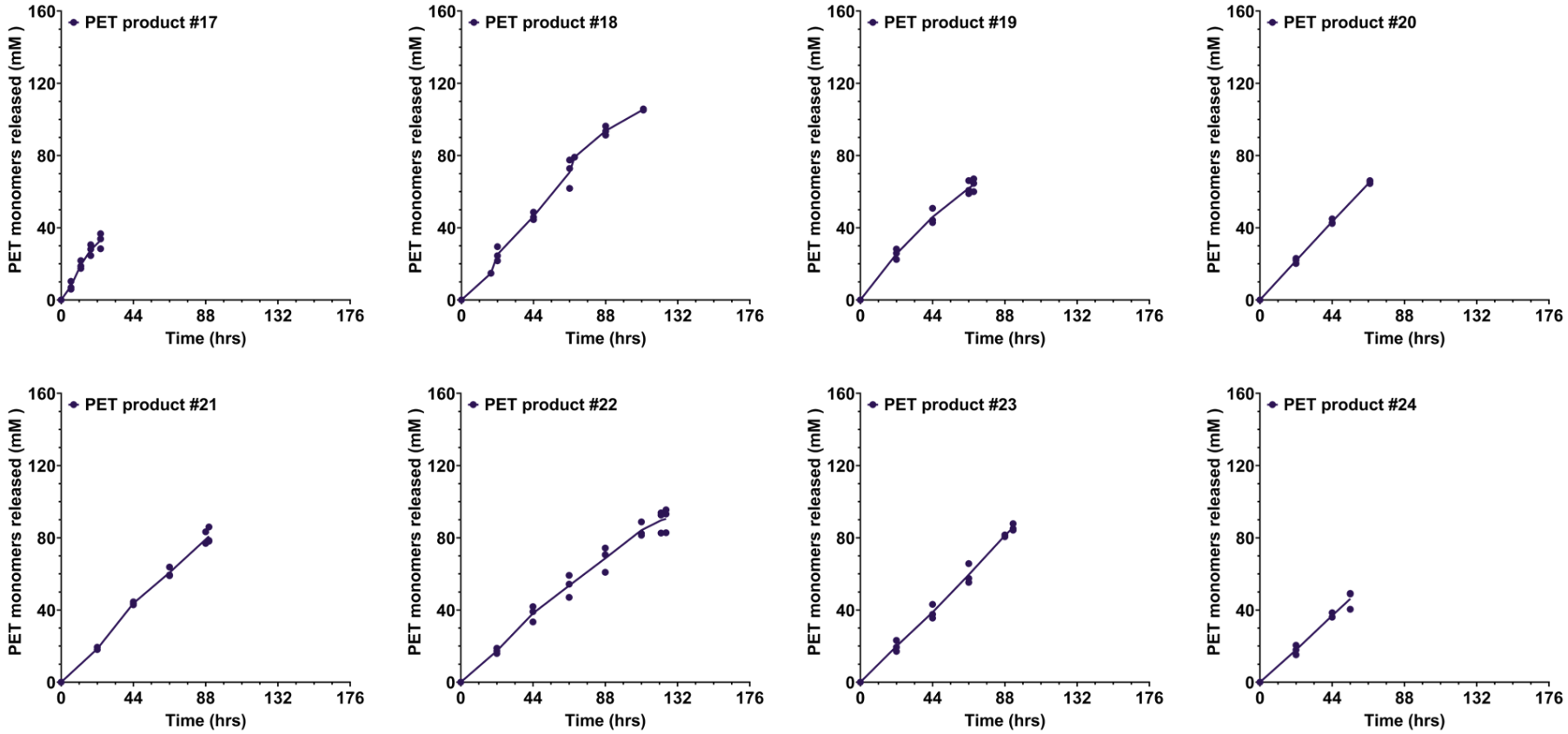
Supplementary Fig. 3 | Sequence alignment of FAST-PETase and selected FAST-PETase homologous enzymes. Mutation N233K (predicted by MutCompute) in PETase corresponded to D250K in Cut190, D238K in LCC, and C238K in ICCM respectively (marked with a red frame). The sequences can be retrieved from the NCBI server with the accession numbers: A0A0K8P6T7 (PETase), BAO42836 (Cut190) and G9BY57 (LCC), whereas the sequence of ICCM can be obtained through modifying the sequence of LCC with substitutions of F243I, D238C, S283C and N246M. The alignment was created with T-Coffee Expresso.



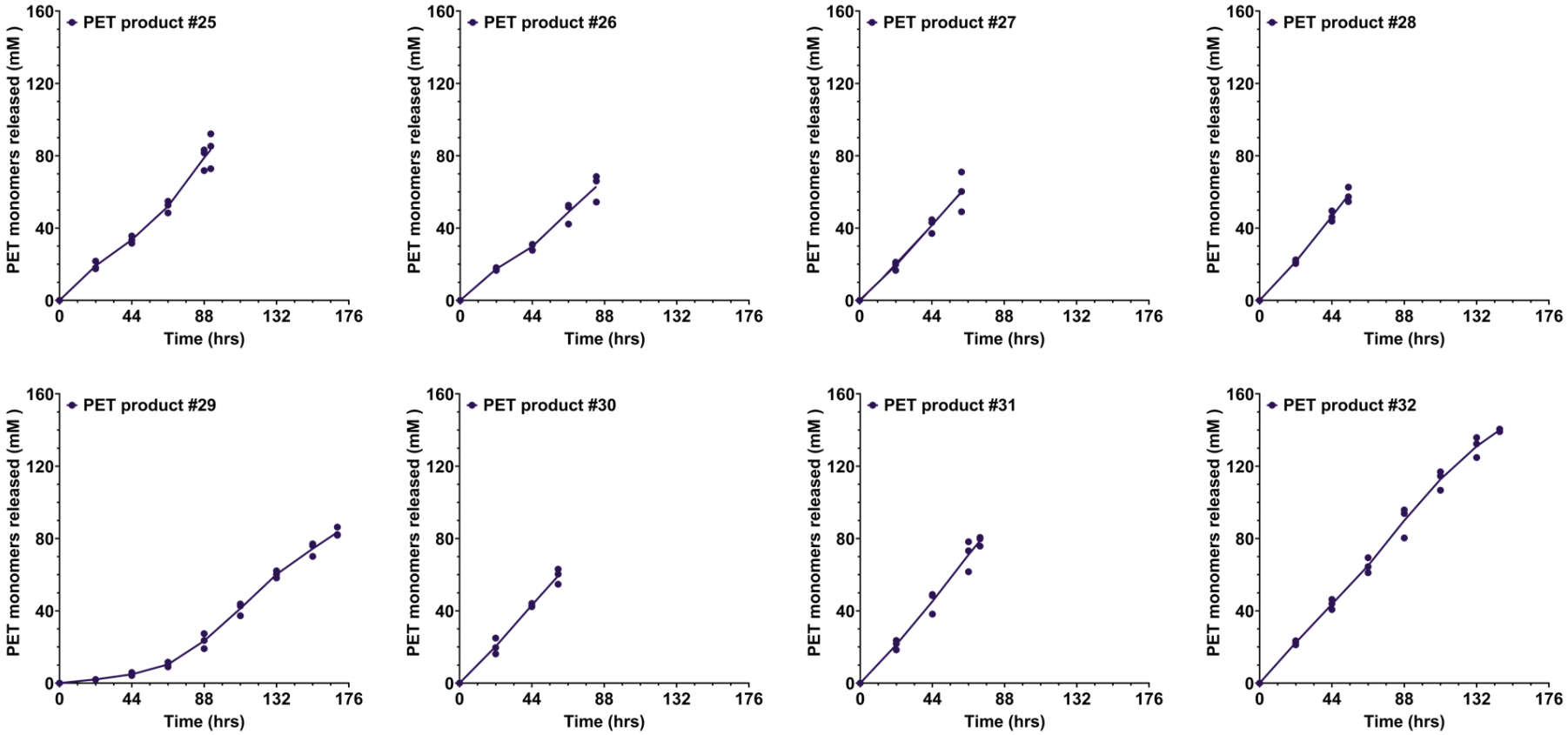
Supplementary Fig. 4 | Time-course of PET monomers released from hydrolyzing the hole-punched pc-PET films of 51 PET products with FAST-PETase. The circular pc-PET films (6 mm in diameter) were hole-punched from 51 different post-consumer plastic products used in the packaging of food, beverages, medications, office supplies, household goods and cosmetics available at local grocery store chains (Walmart, Costco, and HEB). The pc-PET films were hydrolysed by serial treatment with FAST-PETase at 50 ºC until the films were completely degraded (film disappeared). The enzyme solution (200 nM of FAST-PETase in 100mM KH_2PO_4-NaOH (pH 8.0) buffer) was replenished every 22 hours. All measurements were conducted in triplicate (n=3). The circles shown represent the individual numbers.



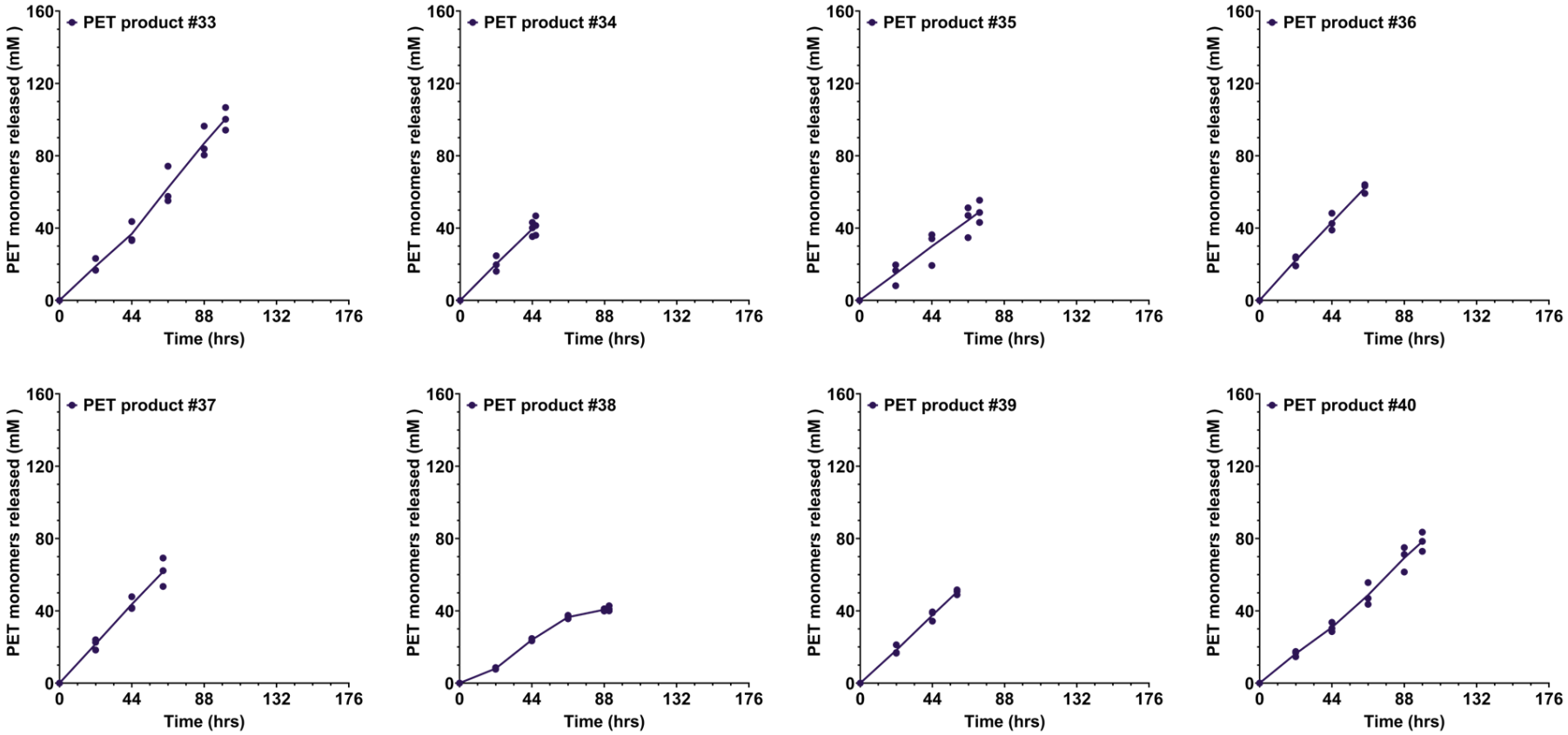
Supplementary Fig. 4 continued



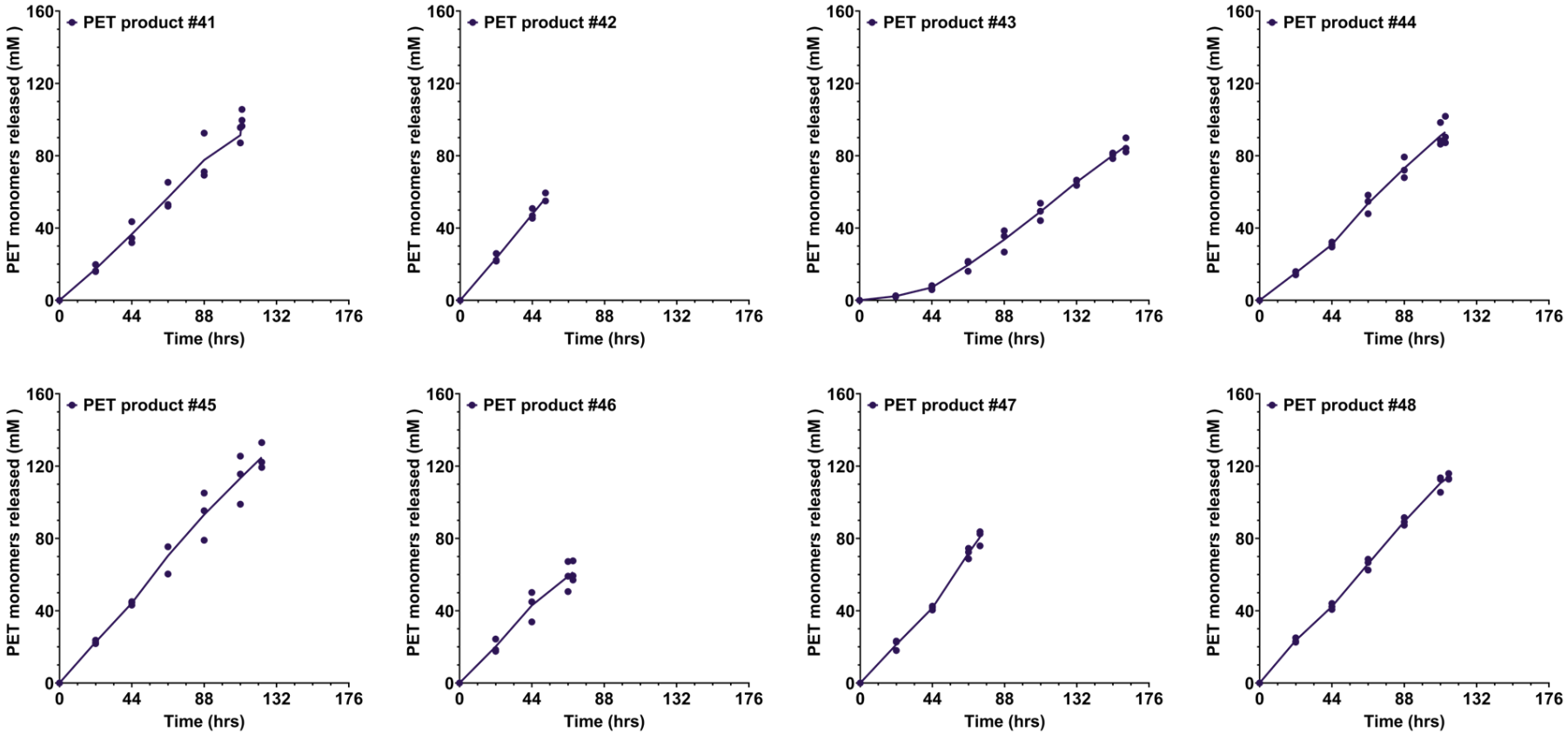
Supplementary Fig. 4 continued



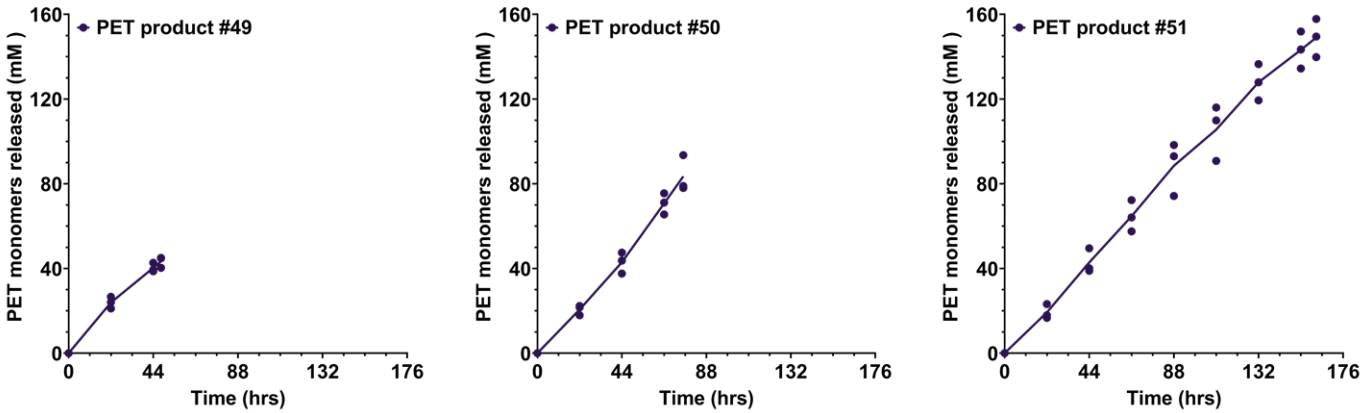
Supplementary Fig. 4 continued



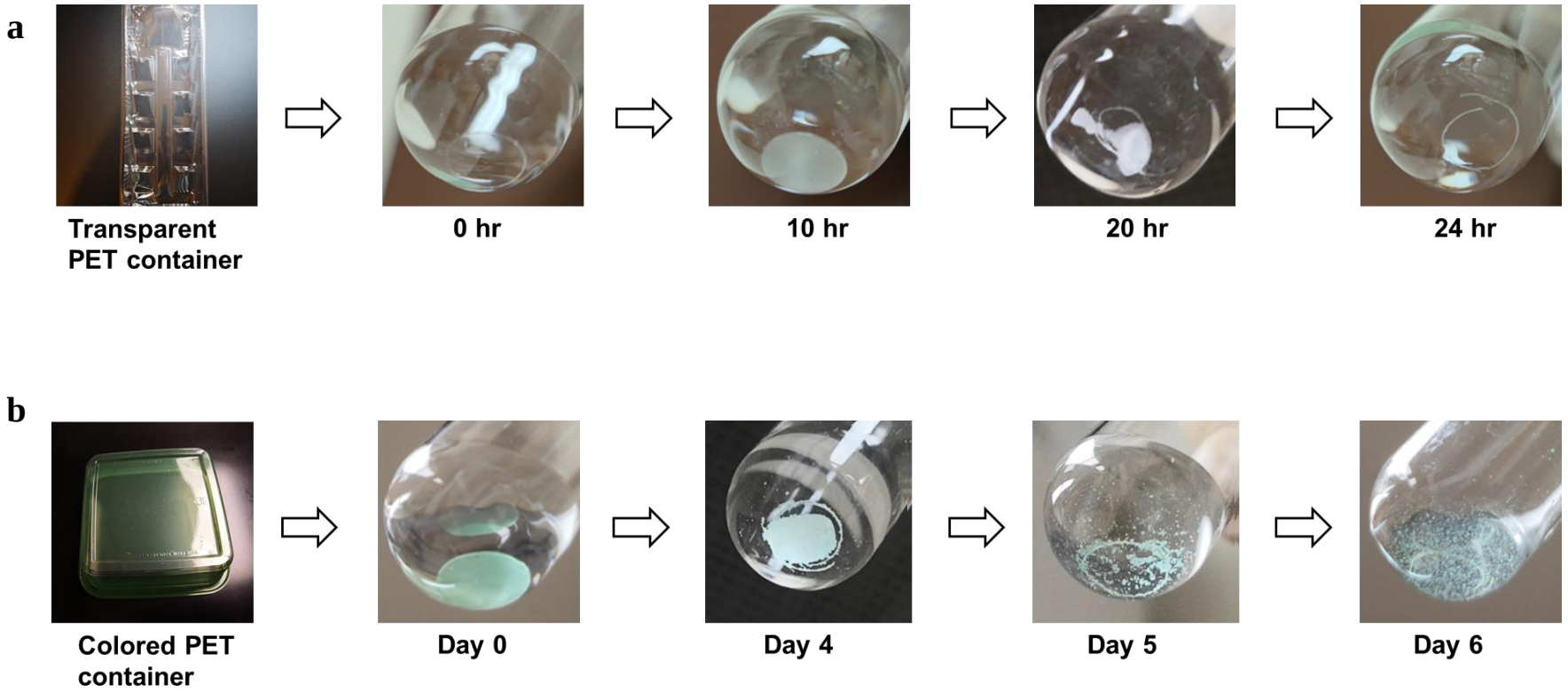
Supplementary Fig. 4 continued



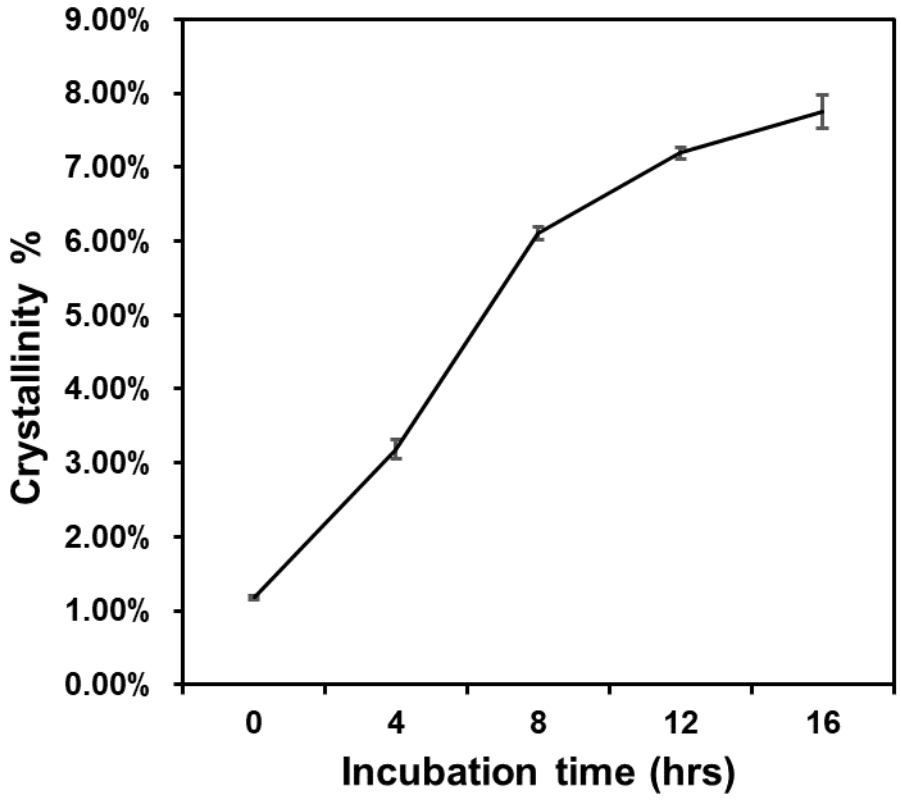
Supplementary Fig. 4 continued



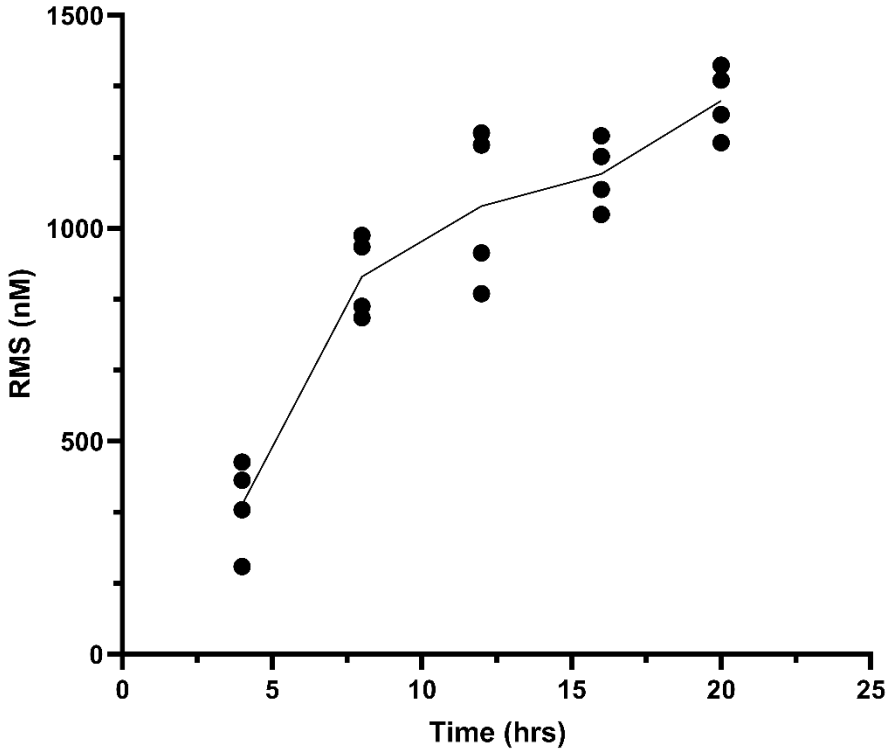
Supplementary Fig. 4 continued



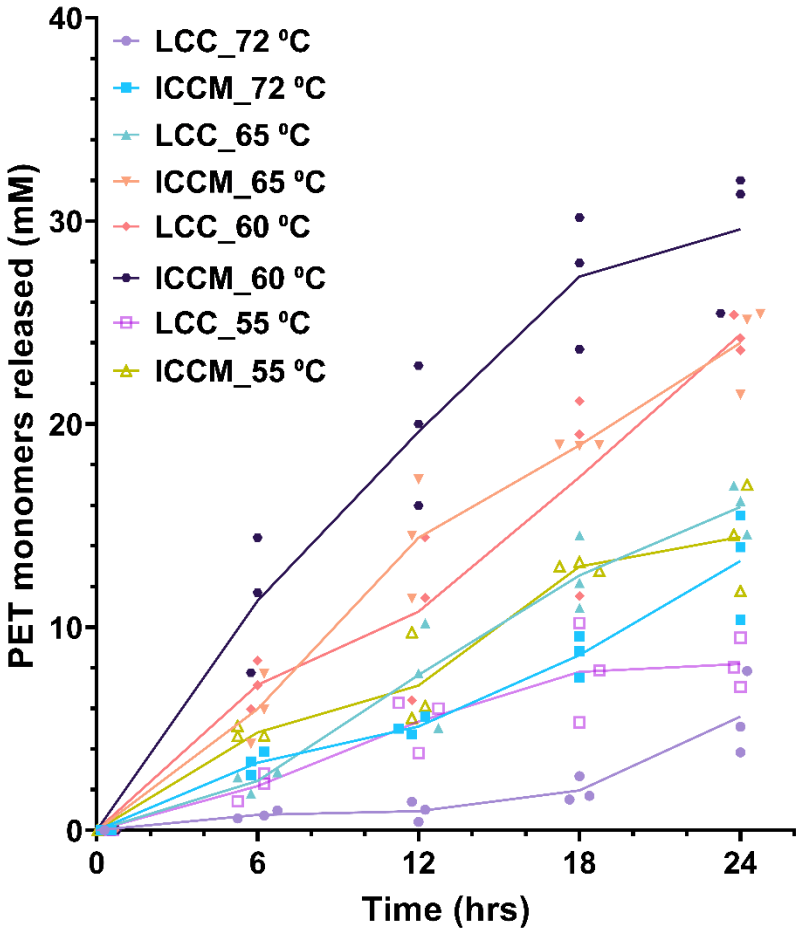
Supplementary Fig. 5 | Stages of degradation of pc-PET films by FAST-PETase. a. The transparent pc-PET film (6 mm in diameter) was completely degraded (only cutting edges of the film remained) after 24 hrs of a single treatment with FAST-PETase at 50 ºC. **b.** The colored pc-PET film (6 mm in diameter) was completely degraded (only cutting edges of the film and some colorants remained) after six days of serial treatment with FAST-PETase at 50 ºC. Enzyme (200 nM) treatment was performed with 100 mM KH_2PO_4-NaOH (pH 8.0) buffer.



Supplementary Fig. 6 | Time-course of crystallinity % of the degraded pc-PET film. The hole-punched PET films from a bean cake PET container were treated with FAST-PETase for 0 hr, 4hr, 8 hr,12 hrs 16 hr in 100 mM KH_2PO_4-NaOH (pH 8.0) buffer at 50 ºC. Crystallinity % of the films was determined by DSC. All measurement were conducted in duplicate (n=2).

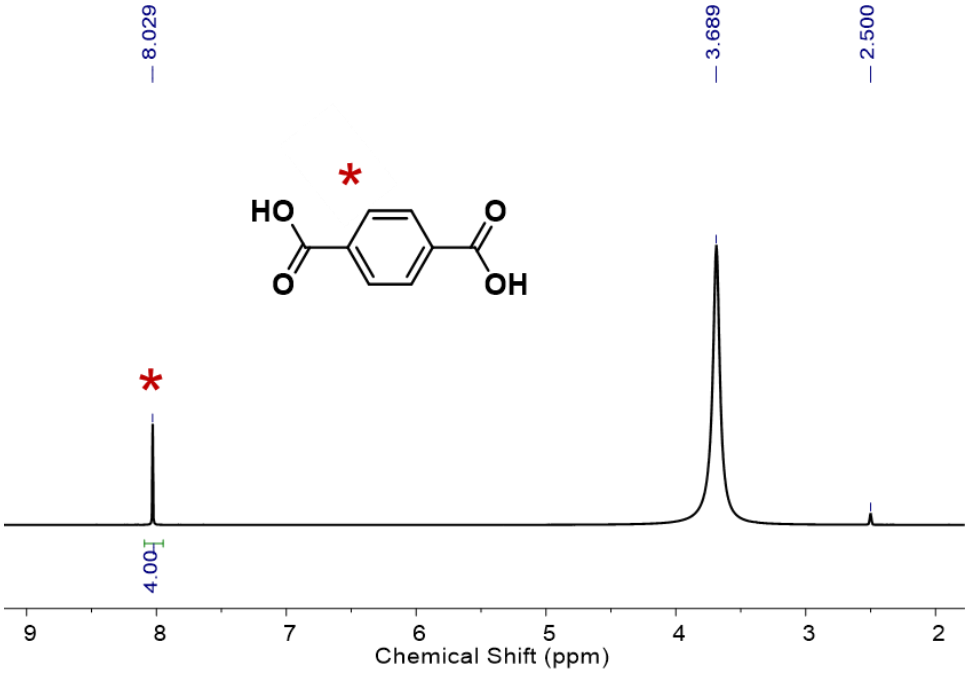


Supplementary Fig. 7 | The surface roughness of the pc-PET films determined by atomic force microscopy. The hole-punched PET films from a bean cake PET container were treated with FAST-PETase for 4 hr, 8 hr, 12 hr, 16 hr and 20 hr in 100mM KH_2PO_4-NaOH (pH 8.0) buffer at 50 ºC. The time-course profile of the surface roughness indicated that longer exposure times with FAST-PETase resulted in higher degree of surface roughness on the pc-PET films. RMS represents root mean square.

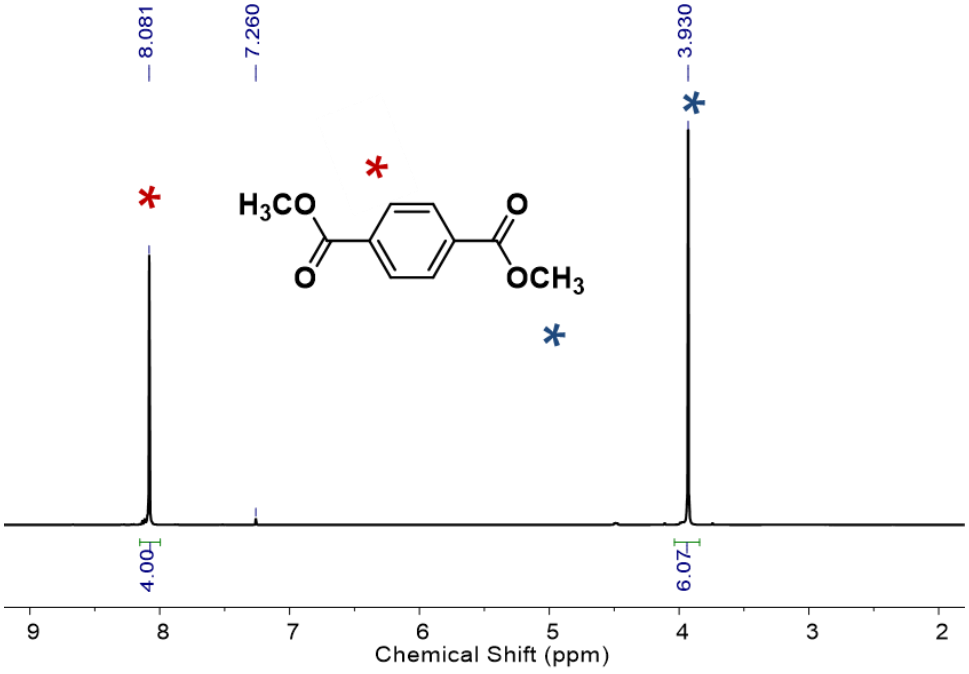


Supplementary Fig. 8 | Time-course of PET-hydrolytic activity of LCC and ICCM at reaction temperatures of 55 ºC, 60 ºC, 65 ºC, and 72 ºC. PET-hydrolytic activity was evaluated by measuring the amount of PET monomers (the sum of TPA and MHET) released from hydrolyzing the pc-PET (Bean cake plastic container) film by the tested PHEs at various time points. 100 mM KH_2PO_4-NaOH (pH 8.0) buffer was used for all reactions shown in this figure. All measurement were conducted in triplicate (n=3).

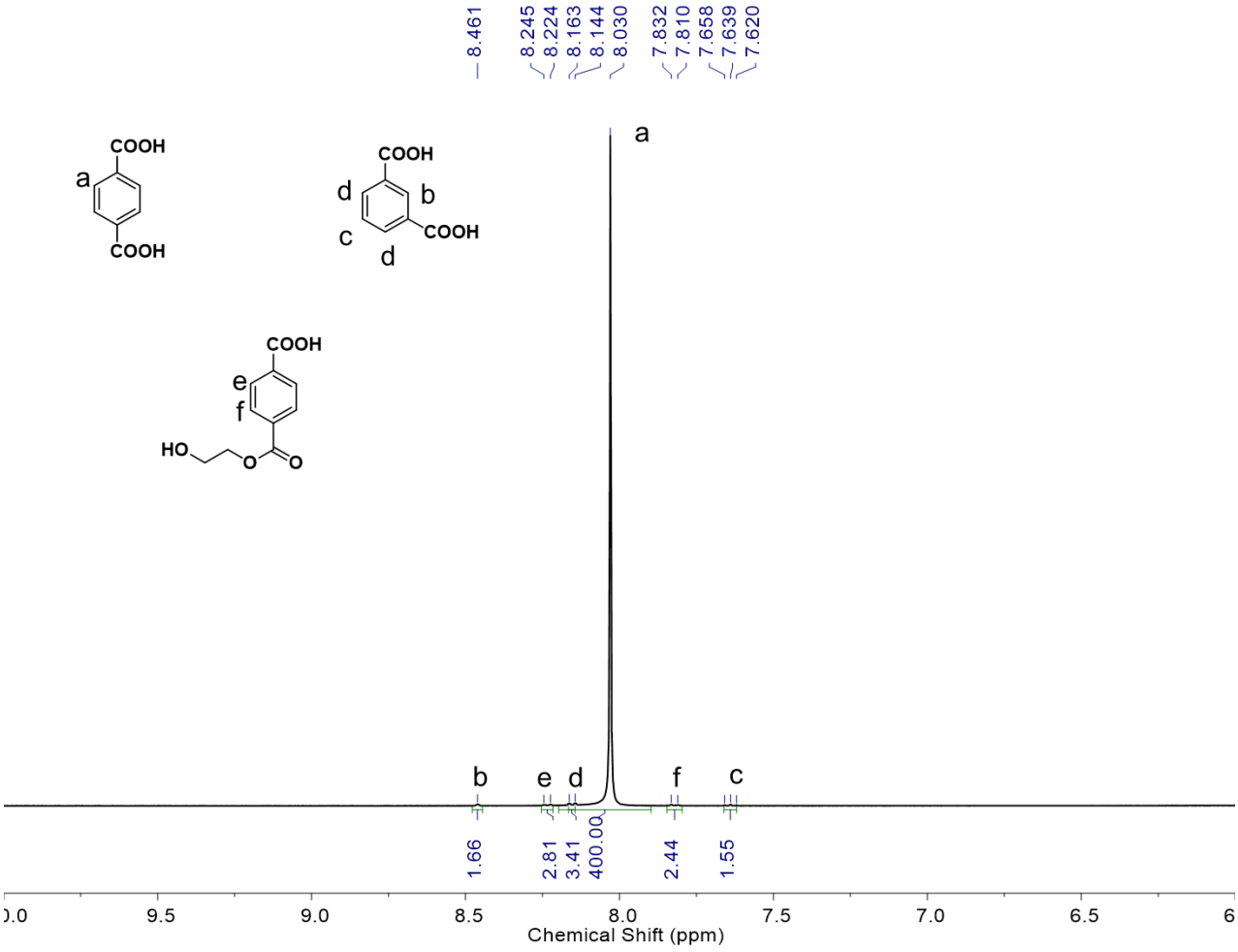
a



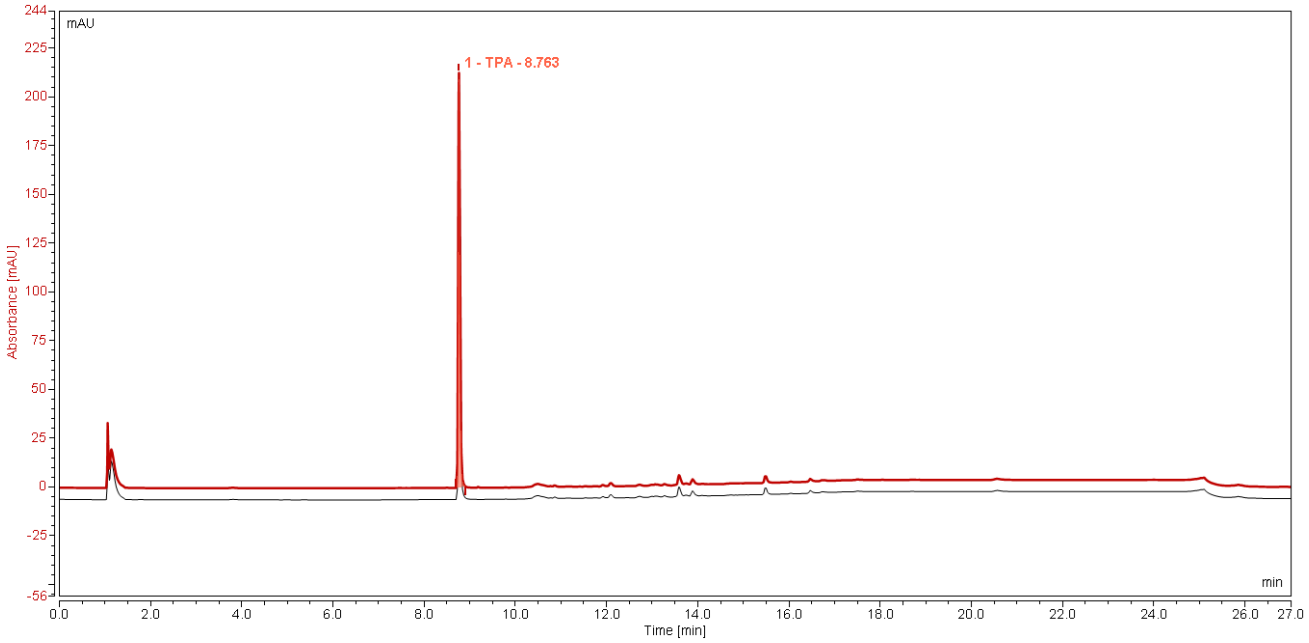
b



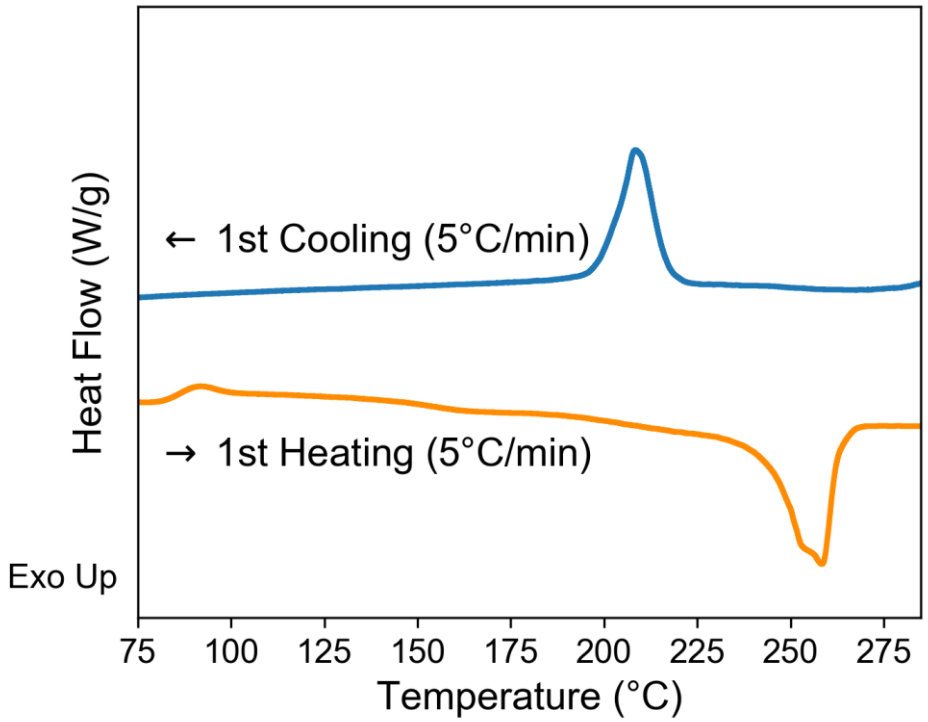
Supplementary Fig 9 | a. ^{1}H NMR (400 MHz, d_6-DMSO) spectrum of TPA recovered from the enzymatic hydrolysate of a colored pc-PET containers depolymerized by FAST-PETase. The peak at 8.029 ppm corresponds to the hydrogen nuclei of the benzene ring. **b. ^{1}H NMR (400 MHz, $CDCl_3$) spectra of DMT synthesized from TPA.** The peak at 8.081 ppm corresponds to the hydrogen nuclei of the benzene ring. The peak at 3.93 ppm corresponds to the hydrogen nuclei of the methyl group.



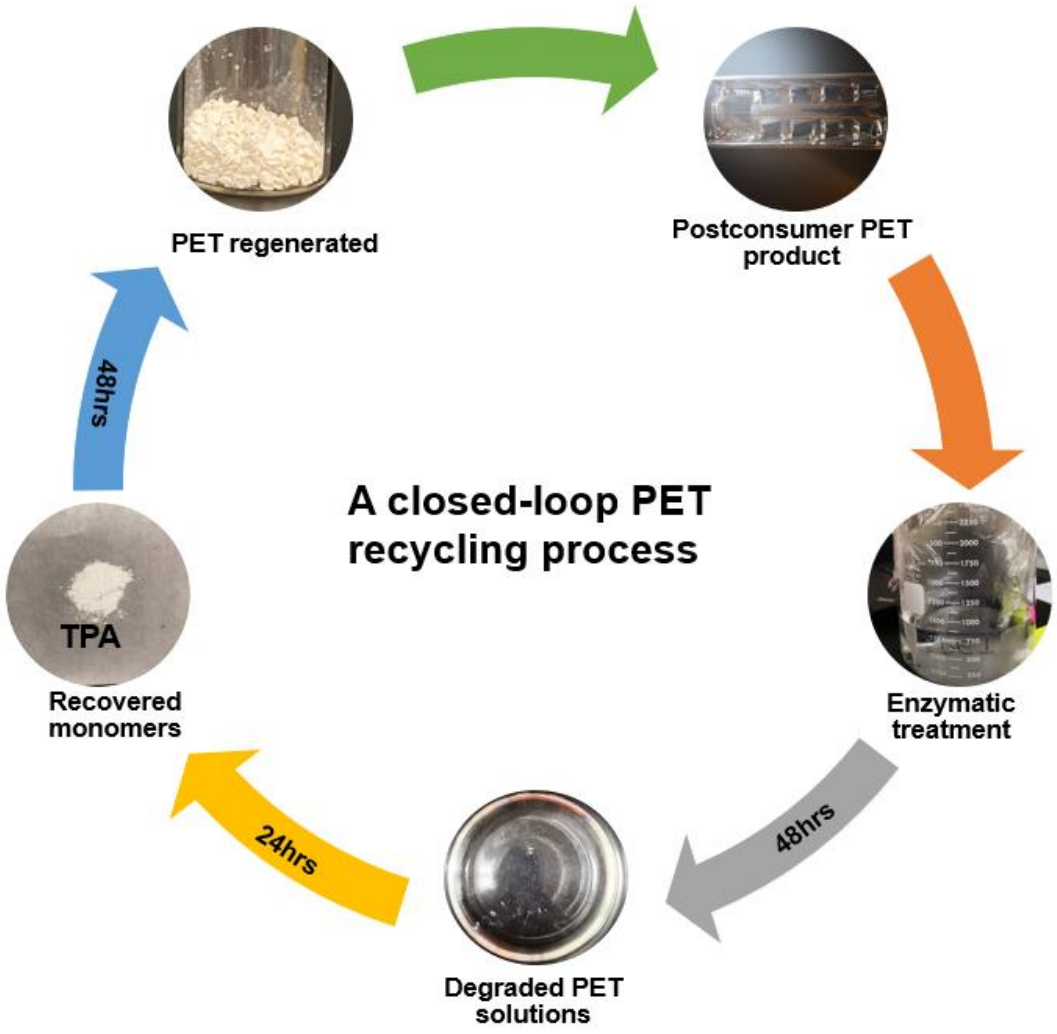
Supplementary Fig. 10 | A zoom-in ^{1}H NMR (400 MHz, d_6-DMSO) spectrum of TPA recovered from degraded PET solutions. The singlet peak at 8.030 ppm corresponds to the hydrogen nuclei of the benzene ring in TPA. The peaks at 8.461 ppm (singlet), 8.163 ppm (doublet), 7.639 ppm (triplet) correspond to the hydrogen nuclei of the benzene ring in isophthalic acid. The peaks at 8.245 ppm (doublet), 7.832 ppm (doublet) correspond to the hydrogen nuclei of the benzene ring in 4-((2-hydroxyethoxy)carbonyl)benzoic acid.



Supplementary Fig. 11 | HPLC chromatogram of purified TPA recovered from the enzymatic hydrolysate of a colored pc-PET containers depolymerized by FAST-PETase. The black line represents the TPA standard (99 % purity), whereas the red line represents the recovered TPA.



Supplementary Fig. 12 | DSC trace of PET regenerated from the degraded solutions. The crystallinity of this regenerated PET is 58.46%. The melting onset is 243.6 °C. The melting peak temperature is 258.4 °C. The glass transition temperature is 84.3 °C.



Supplementary Fig. **13 | A closed-loop PET recycling process.** Demonstration of a closed-loop process for enzymatically degrading and then regenerating PET in the course of several days.

Reference

1. Fujita, M. *et al.* Cloning and nucleotide sequence of the gene (amyP) for maltotetraose-forming amylase from Pseudomonas stutzeri MO-19. *J. Bacteriol.* (1989). doi:10.1128/jb.171.3.1333-1339.1989

2. Riesselman, A. J., Ingraham, J. B. & Marks, D. S. Deep generative models of genetic variation capture the effects of mutations. *Nat. Methods* **15**, 816–822 (2018).

3. Shin, J.-E. *et al.* Protein design and variant prediction using autoregressive generative models. *Nat. Commun.* **12**, 2403 (2021).

4. Shroff, R. *et al.* Discovery of novel gain-of-function mutations guided by structure-based deep learning. *ACS Synth. Biol.* (2020). doi:10.1021/acssynbio.0c00345

5. Paik, I. *et al.* Improved Bst DNA Polymerase Variants Derived via a Machine Learning Approach. *Biochemistry* (2021). doi:10.1021/acs.biochem.1c00451

6. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

7. Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).

8. McNutt, A. T. *et al.* GNINA 1.0: molecular docking with deep learning. *J. Cheminform.* **13**, 43 (2021).

9. van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **428**, 720–725 (2016).