

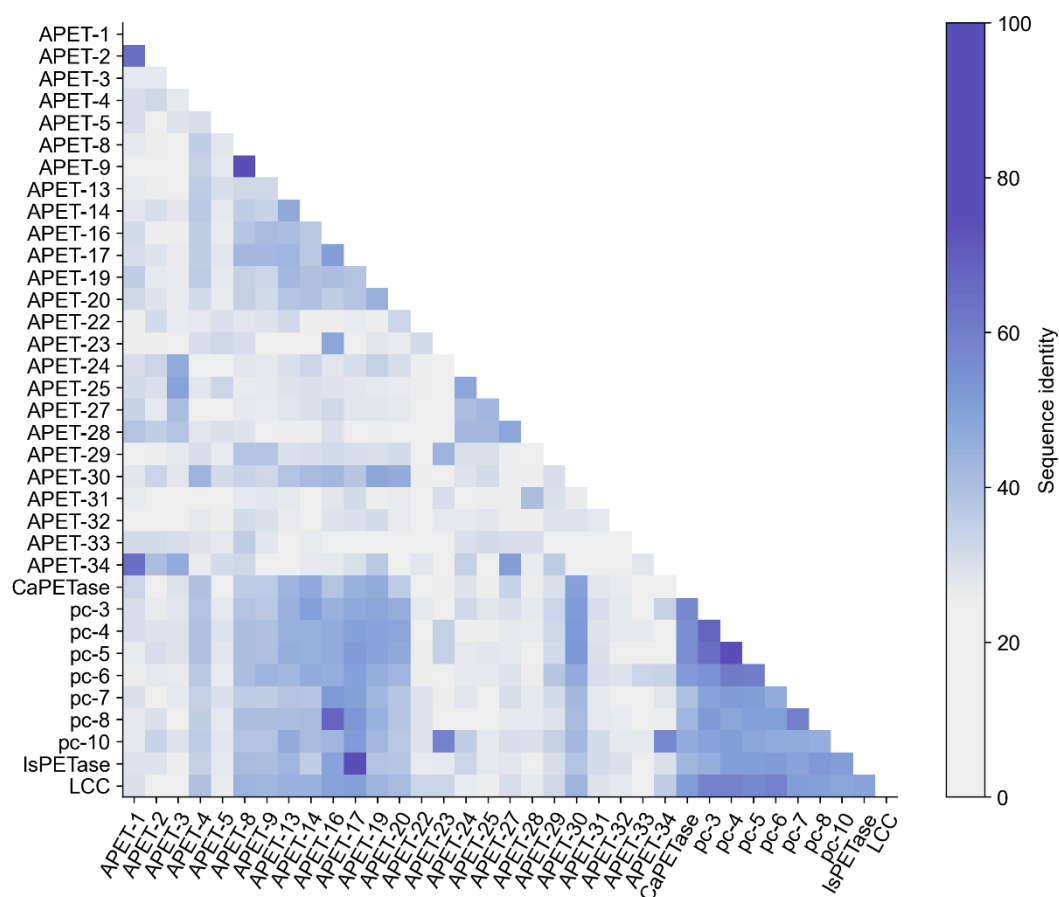
Supplementary Information for:

Harnessing Protein Language Model for Structure-Based Discovery of

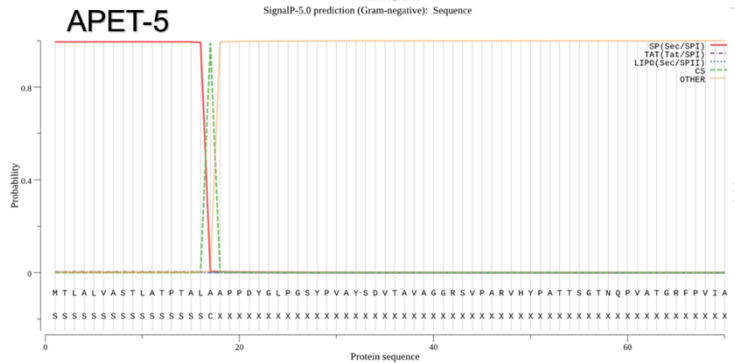
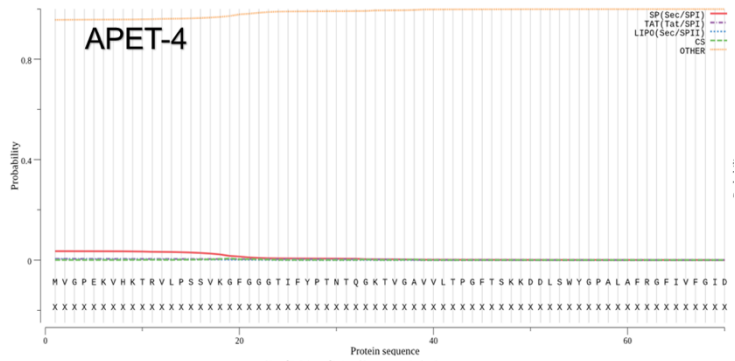
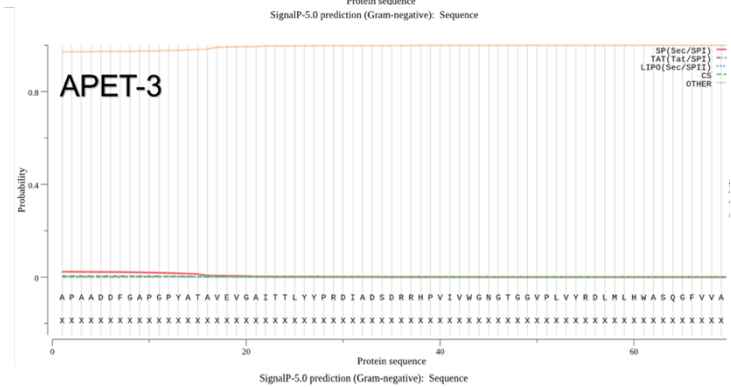
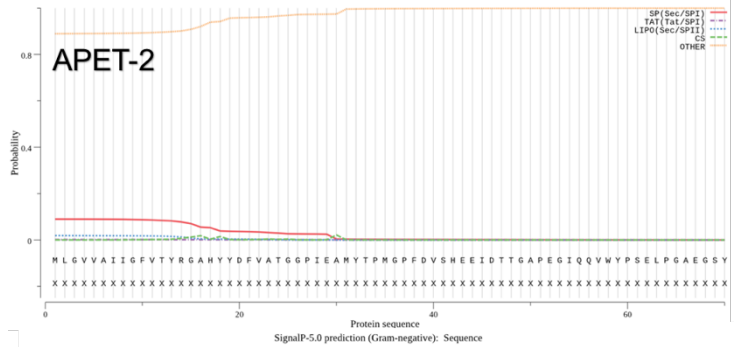
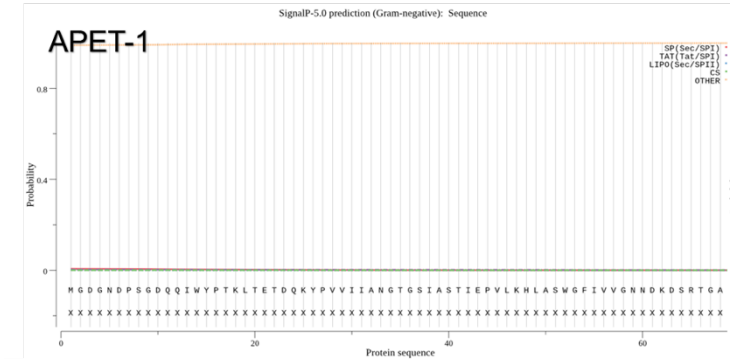
Highly Efficient and Robust PET Hydrolases

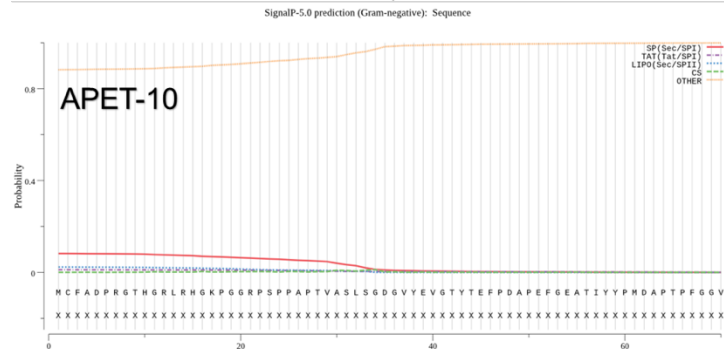
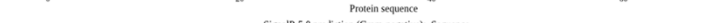
Banghao Wu^{1,2,4†}, Bozitao Zhong^{1,2†}, Lirong Zheng^{2,5*}, Runye Huang^{1,2,4}, Shifeng Jiang¹, Mingchen Li^{2,3}, Pan Tan^{1,2,3*} and Liang Hong^{1,2,3,4*}

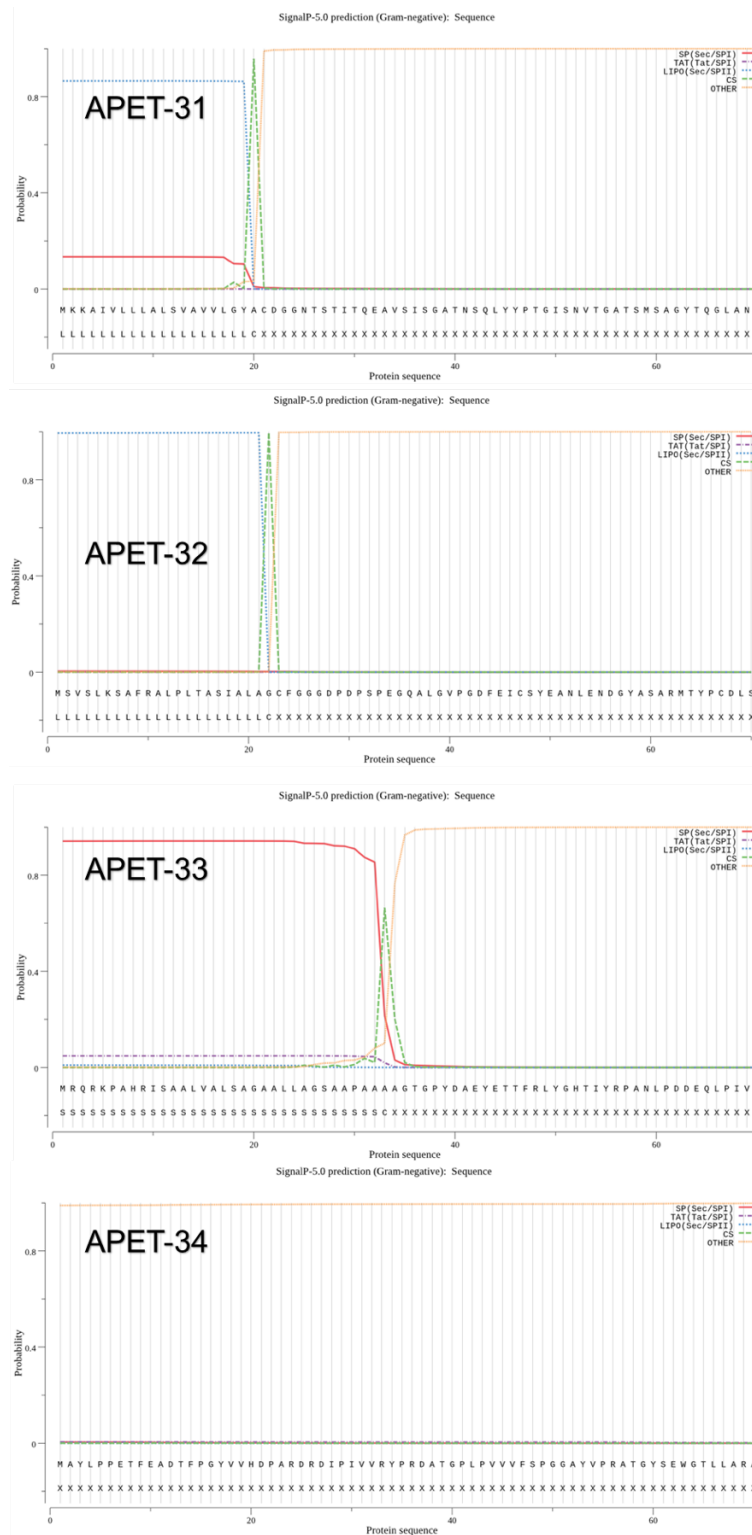
1. School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China
2. Shanghai National Center for Applied Mathematics (SJTU Center) & Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200240, China
3. Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China
4. Zhang Jiang Institute for Advanced Study, Shanghai Jiao Tong University, Shanghai 201203, China.
5. Department of Cell and Developmental Biology & Michigan Neuroscience Institute, University of Michigan Medical School, Ann Arbor, Michigan, 48104, USA



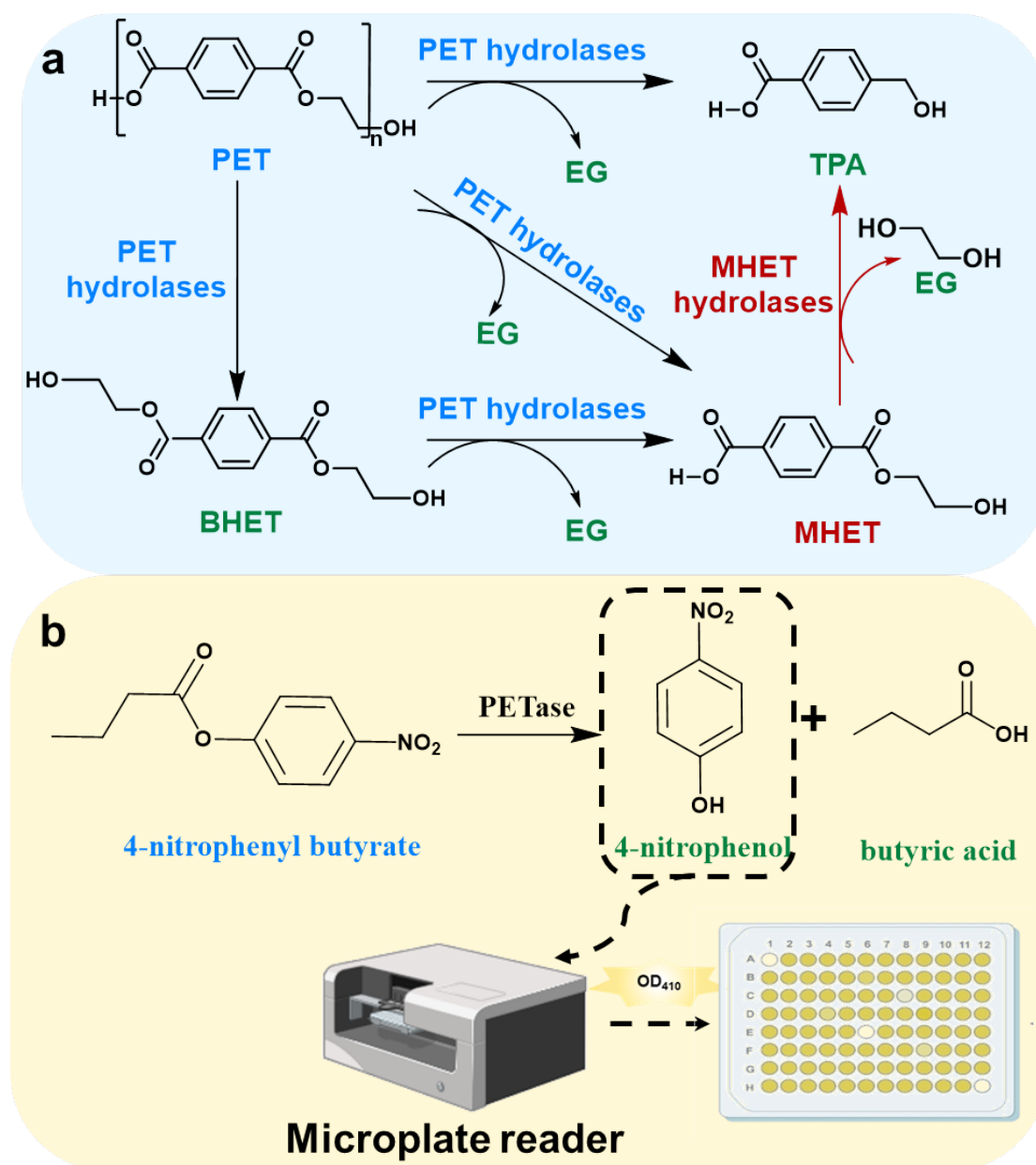
Supplementary Figure 1. The 34 candidate proteins identified as APET-1 to APET-34 were compared for sequence similarity with known PETases. Except for APET-17, which exhibited over 60% similarity, the sequence similarity of the other candidates to the template *IsPETase* ranged between 30% and 40%. Additionally, all candidates showed sequence similarities of 20% to 40% with LCC, indicating a relatively low level of sequence similarity overall. This approach enables the discovery of a substantial number of proteins with low sequence similarity, broadening the potential for enzyme engineering.



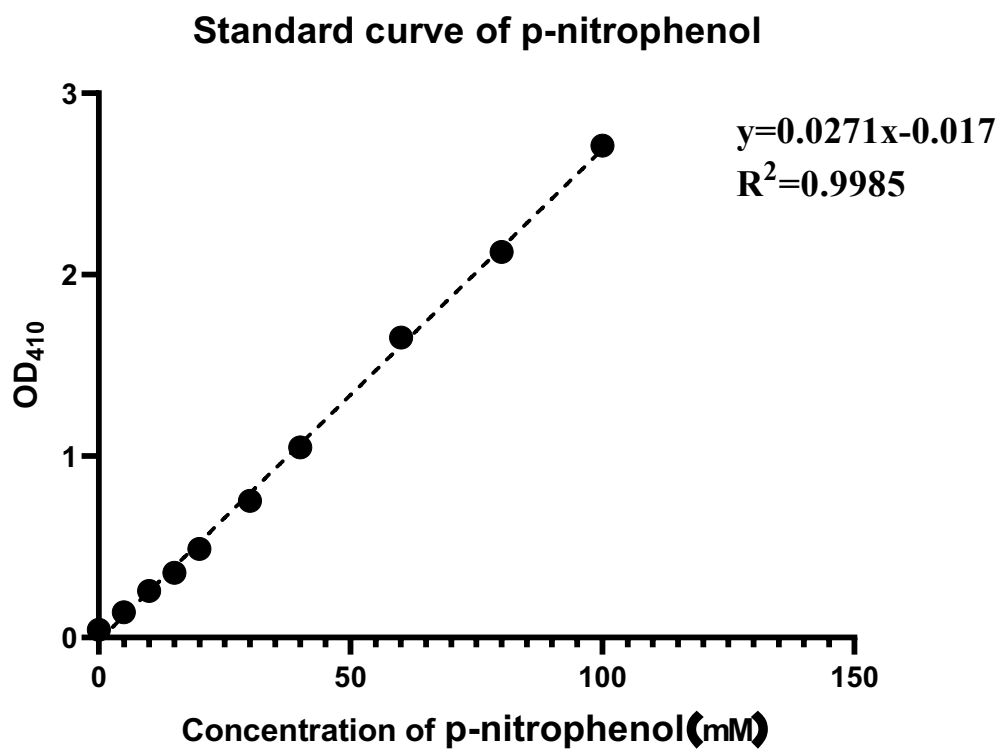




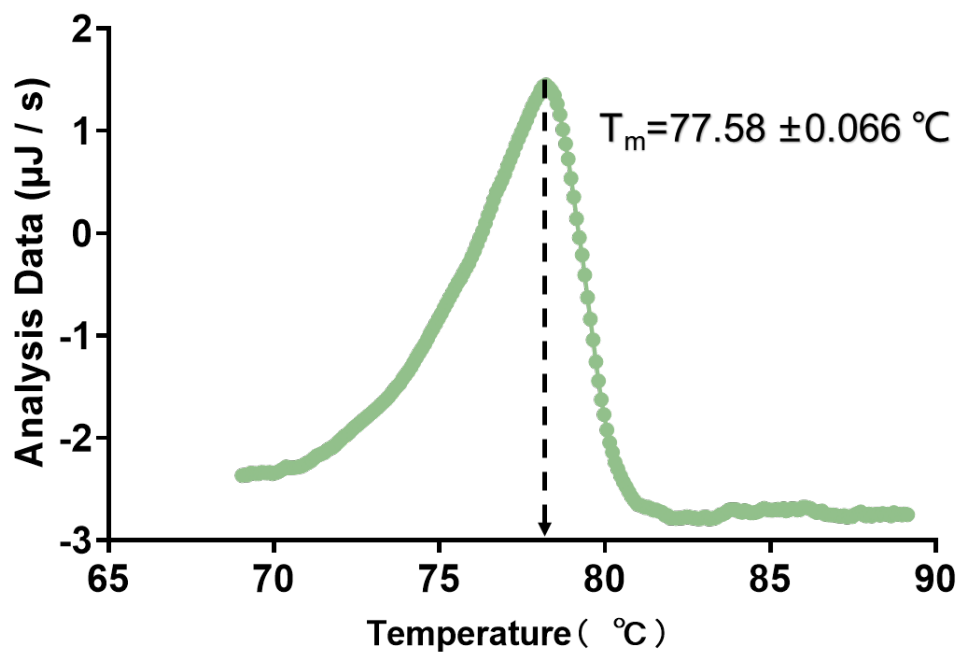
Supplementary Figure 2. The likelihood of presence of a signal peptide in APET-1-APET-34 sequences, as predicted by SignalP5.0¹. APET-5, APET-7, APET11-21, APET-24-29 and APET-33 are predicted to contain Sec signal peptides cleavable by SPase (Sec/SPI). APET-8, APET-9, APET- 31 and APET- 32 are predicted to contain Sec (Lipoprotein) signal peptides cleavable by SPase (Sec/SPII). APET-23 is predicted to contain Tat signal peptides cleavable by SPase (Sec/SPI).



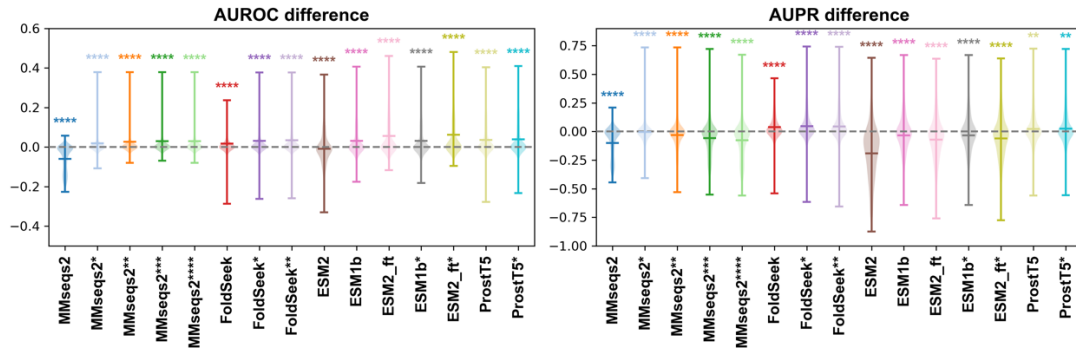
Supplementary Figure 3. The schematic for the initial screening of PET hydrolase candidates using the substrate *p*-nitrophenyl butyrate (*p*NPB) is illustrated as follows². (a) The hydrolysis process of PET polymers; (b) shows that *p*NPB can be hydrolyzed by PETase, resulting in the formation of 4-nitrophenol and butyric acid. Notably, 4-nitrophenol exhibits a peak absorbance at 410 nm. By monitoring changes in absorbance at this wavelength, we can assess the ester bond hydrolysis activity of the candidate proteins, which serves as an indirect indication of their potential PET hydrolytic capabilities.



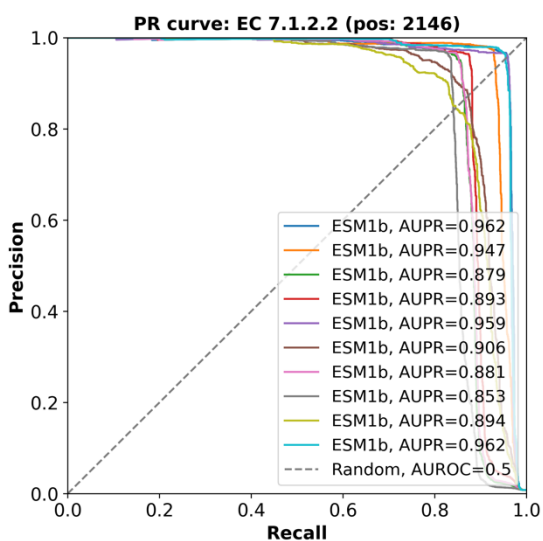
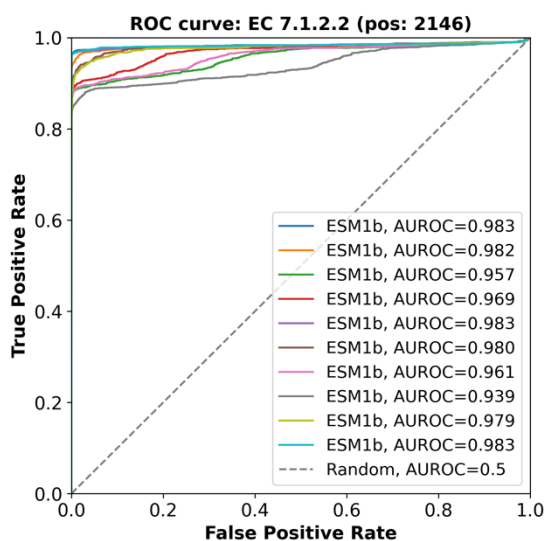
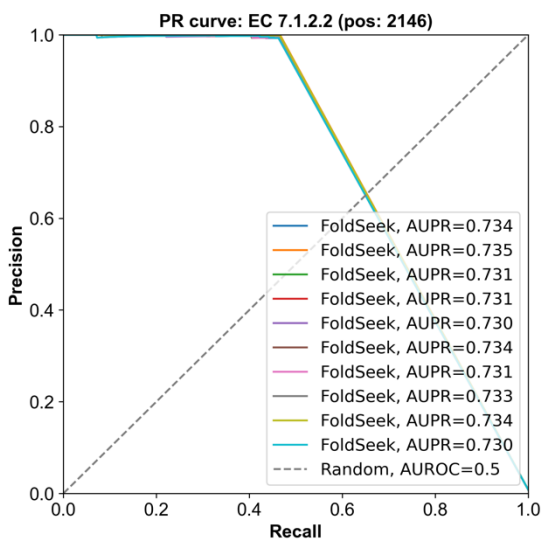
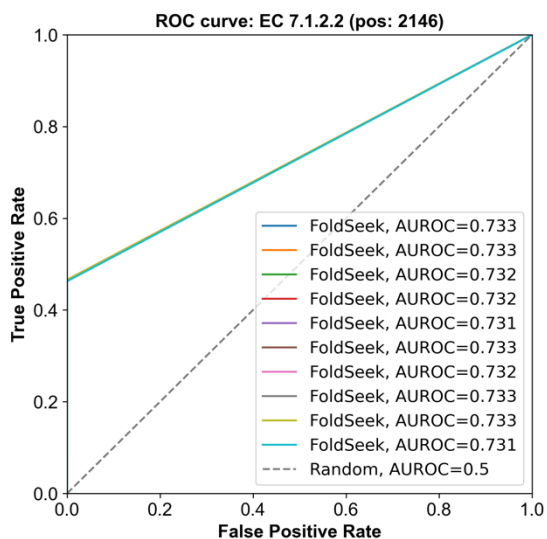
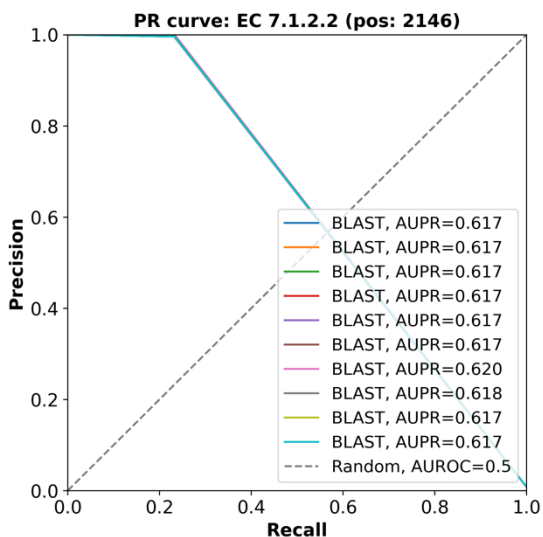
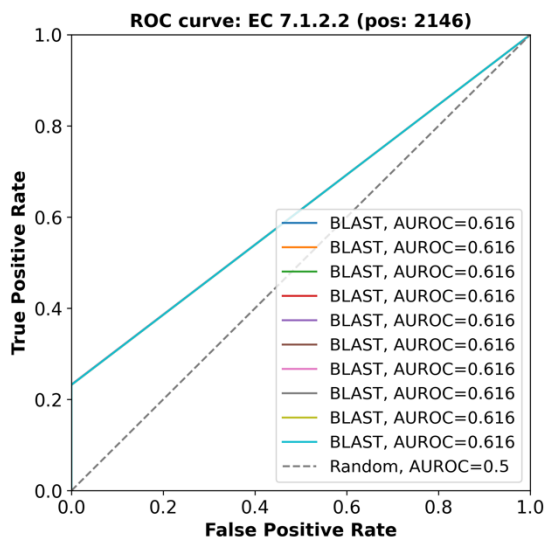
Supplementary Figure 4. The standard curve of p-nitrophenol, which is employed for the quantitative assessment of the ester bond hydrolytic capabilities of the candidate PETases². By utilizing this methodology, we can effectively evaluate the hydrolytic activity of each candidate, providing insights into their potential efficiency in degrading PET substrates.

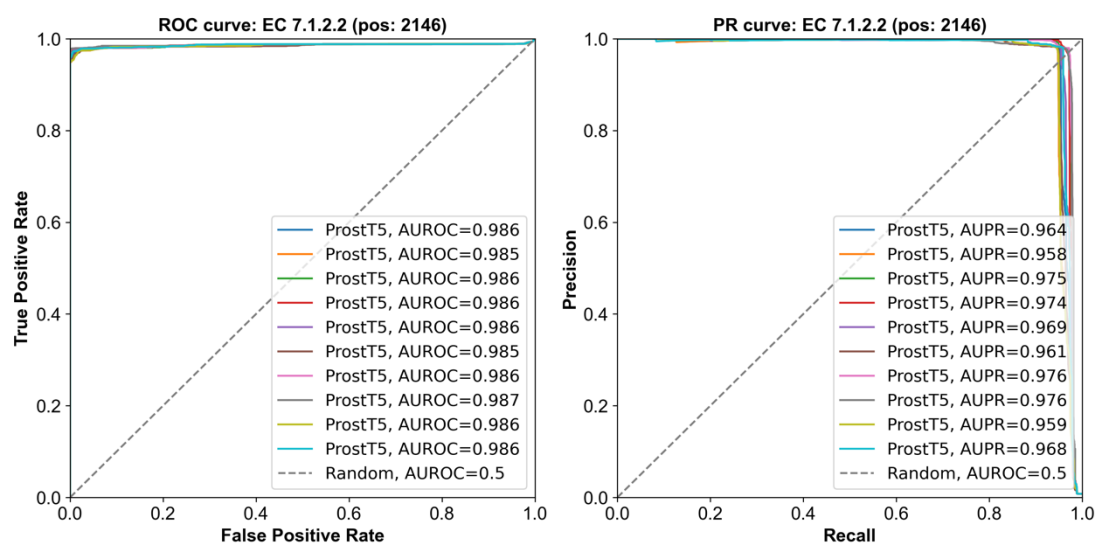


Supplementary Figure 5. The melting temperature (T_m) of *KbPETase* was re-evaluated using nano DSC. The results indicated a T_m value of $77.58 \pm 0.066 \text{ } ^\circ\text{C}$, which closely aligns with the value obtained through DSF, further validating the high thermal stability of *KbPETase*.

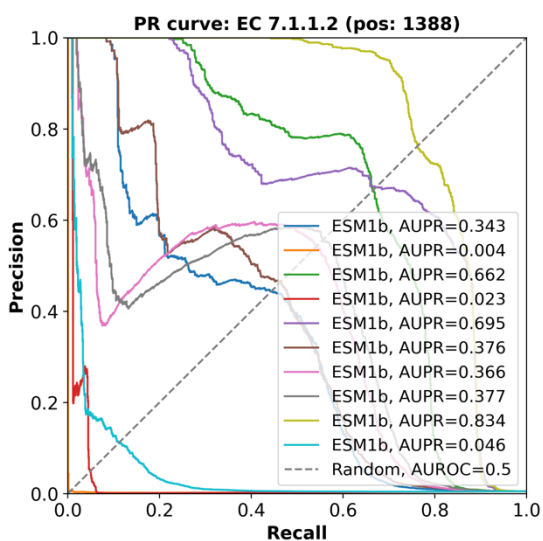
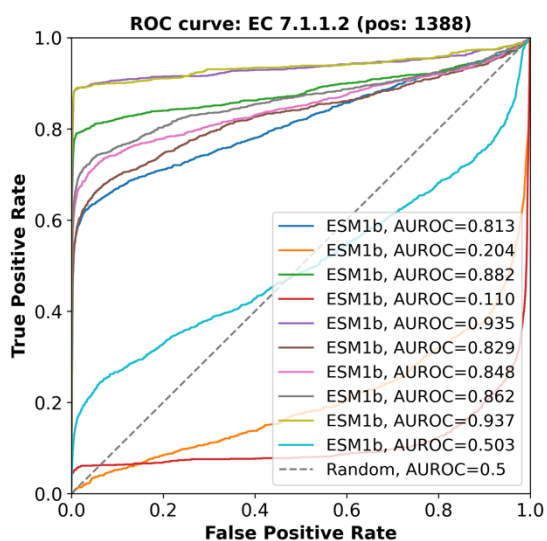
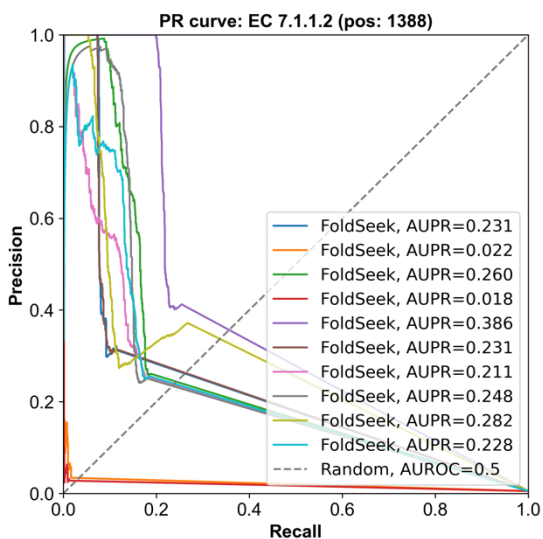
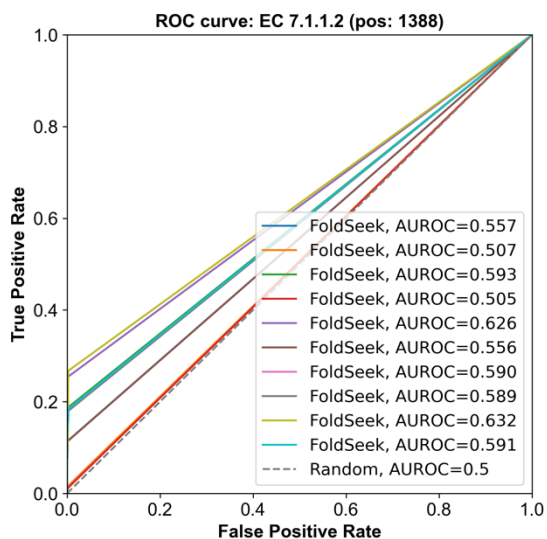
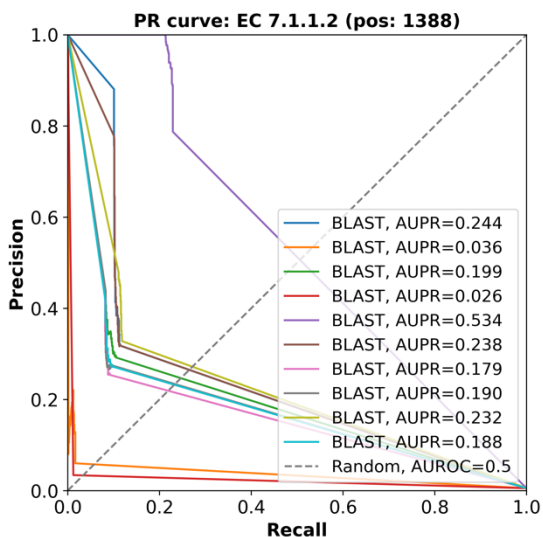
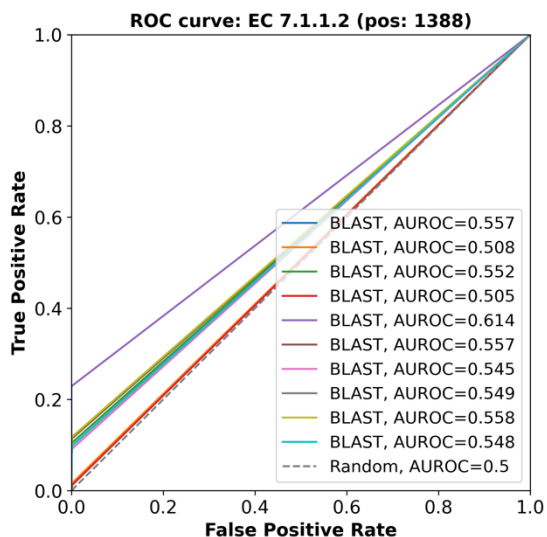


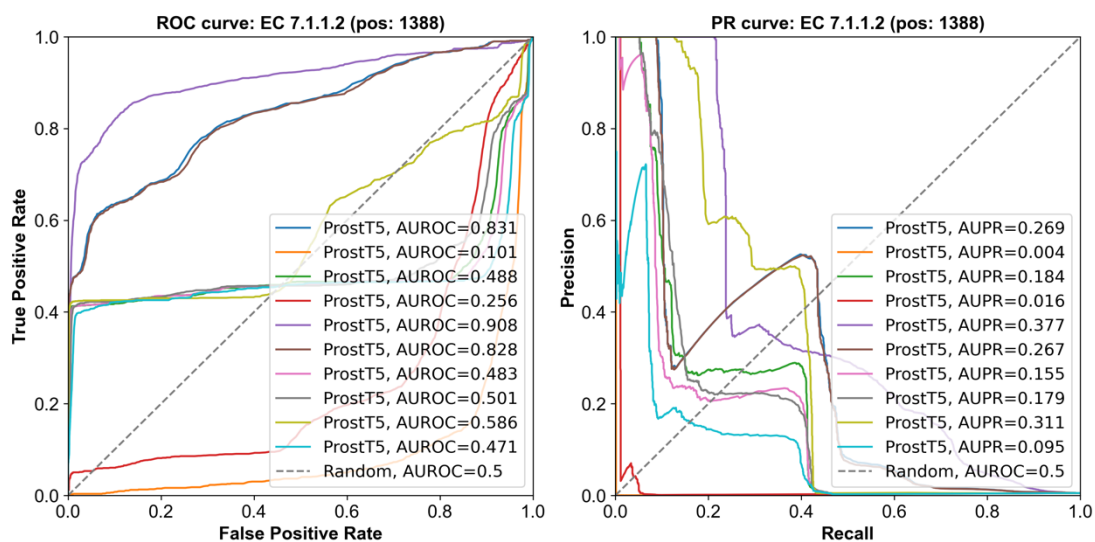
Supplementary Figure 6. Comparison of different models towards BLAST on AUROC and AUPR difference for protein discovery. Different methods are tested including MMseqs2 (various maxseq), FoldSeek (various maxseq), ESM2 (Euclidean distance or cosine similarity), ESM1b (Euclidean distance or cosine similarity), fine-tuned ESM2 (Euclidean distance or cosine similarity), and ProST5 (Euclidean distance or cosine similarity). For language models, * are marked for cosine similarity. MMseqs2 including using different max sequence setting including 1000 (default), 2500 (*), 5000 (**), 10000 (**), 20000 (**). FoldSeek including 3 maxseq settings including 1000 (default), 2500 (*) and 5000 (**). Two sided Wilcoxon signed-rank test are employed to compute the significance of the difference if AUROC/AUPR between the target method and BLAST. The asterisks above the violin plot indicate the p-value for two-sided Wilcoxon sum rank test, ranging from 0.01~0.05 (*), 0.001~0.01 (**), 0.0001~0.001 (***), and <0.0001 (****).



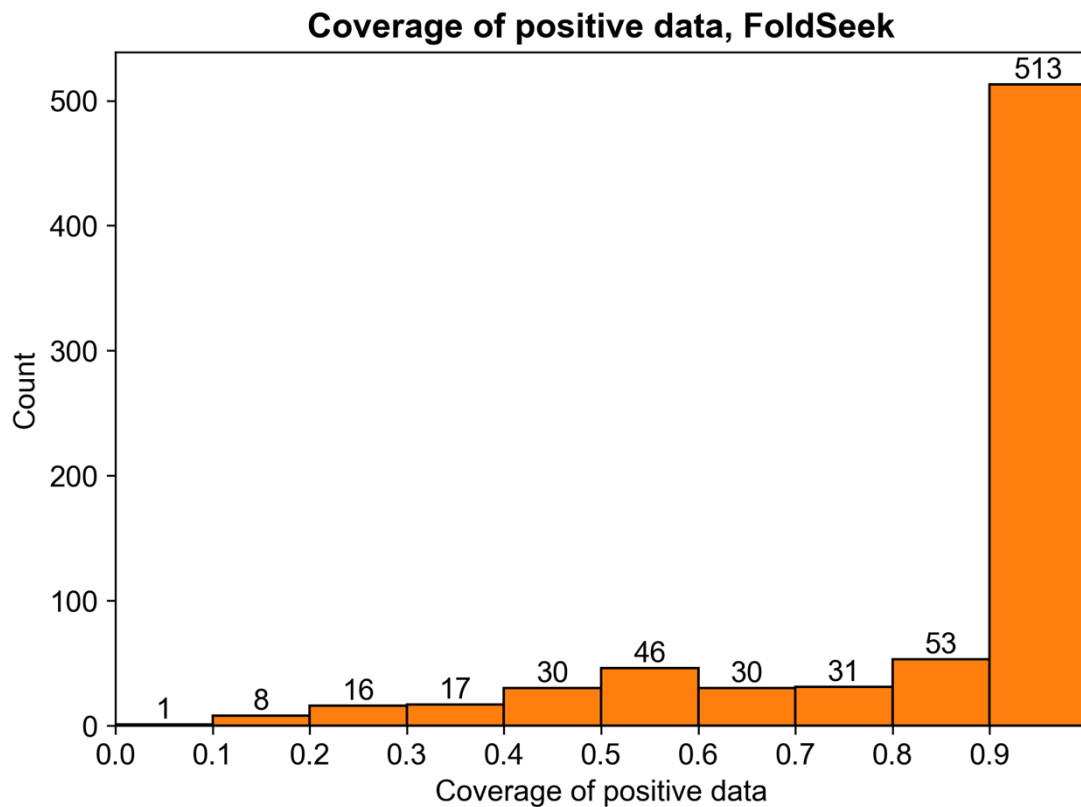
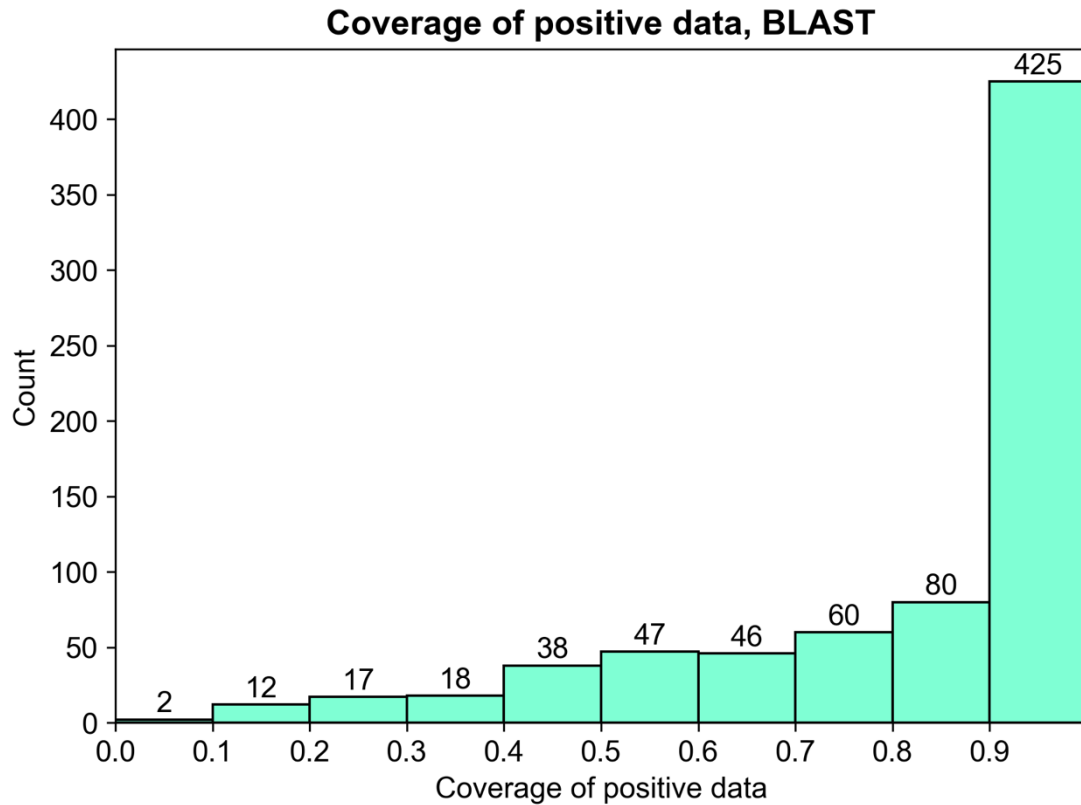


Supplementary Figure 7. The ROC curve and PR curve when evaluate the BLAST, FoldSeek, ESM-1b and ProstT5 ability to find enzymes with same EC number 7.1.2.2. Curves with 10 colors indicate the 10 randomly selected starting points.





Supplementary Figure 8. The ROC curve and PR curve when evaluate the BLAST, FoldSeek, ESM-1b and ProstT5 ability to find enzymes with same EC number 7.1.1.2. Curves with 10 colors indicate the 10 randomly selected starting points.



Supplementary Figure 9. Comparing the coverage of BLAST and FoldSeek founded positive candidates as a fraction of all positive candidates (true positive rate, recall). BLAST and FoldSeek both selected E-value $> 10e-3$ as cutoff. X-axis marked the recall range; y axis marked the count of EC number.

Supplementary Table 1. Nucleotide sequences of APET1-34.

Genes	Sequences
APET-1	catatgggtgacggcaatgatccgtctggtagaccagcagatctggtagccgaccaaactgaccgaaaccgatcag aaatacccggttggtatcatcgtaacggtaccgggtctatcgctctaccatcgaaccgggtctgaaacacctggcg agctgggggttcatcgtggtggcaacaacgacaaagacagccgtactggcgcgagctcgggaagcgagcctgga tttcatcgttctcgtaacgatgacaccaacagcatctcaacggccatggtgatgttgacaacatcggtatagcgggt cacagccagggtggcgtgggtgcgatcaacgctgcgaccaaacagccgtctgctgataaatactctgcgctgttc gttgcgtccccgaccggtgaatactgggggtaccaacccggatttcggcccggaatggctctacaacccgctaaa ctgtctatcccaccttcgcgatcgcgggtaccggccggtcgtatgcgggtaccgcgaccgatataacgcaacc aaagggtcagggtcagcccgtgtggagcctgcaggacaccttaatgcgatcccggtgatgttgcgaaaac cgttgctcgtcgtgtggcgcggtatcatggcgataccctgaccagcgggtccggttacatgaccggttggttcgca tactgggtgcagggttaaccagaaagcggcgcggttctctggcgatcgtccagaaatcggaacaacccgaa atggcaggatctgcacatcaaaatcggttaactaagcggccgc
APET-2	catatgcactactacgatttcgttgcgaccggtggcccgatcgaagcaatgtacaccccgatgggtccgttcgatgt ttccacgaagaaatcgacaccaccggcgcgccggaaggtatccagcaggtgtggtaccgctgtaactgccgg gtgcggaaggtcctaccgggtggtggtgatcgcgaacggcaccggcgttcggcggtccaaaaccgaaccggc gctgcgtcacctggcgtcctggggcttcatcgttatcggcaacgaagatgaaaacagccgcaccggcgaaagca gcgtggcgagcctgcagttcatcctggacagcaacgatgacccgcacagcaccttcacgatcacgtggacacc agcaacgttggcatcgcgggtcactctcagggtggcgtggcgcgatcaacgcggcgaccgcgctccgggc gggtgatcgcttactgcgctgtttgccgtagcccgactggctgttactgggggcaggatgccgtgttcgggtccgg agtggctctacgatgtaccggtctggcaattccgtccttctgatcggcggtaccgggtccttcgatcggggcacc gcgaccgatatctccgccaccgacgggtcagggtatcagcccgtgtggtctctggaagataccttctgtacatcc cggattctgttaccaaagtattgcactgcgtgatgacaccgaccacagcgatacgtgaccaccggctctggcta catgaccgatggttcgttactgggtgcagcagcaccaggacgctgctggtgcattcaccgggtaccacgccgga aattaccctgaacgacaactggaaaaacgtgcgtatcgggtggaataagcggccgc
APET-3	catatggcgccggcgccggatgatttcggcgcgccgggtccgtatgcgaccgcggtgaagtgggtgcgatcac caccctgtactaccgcgtgatacgtgatagcgaccgctcgtcaccgggtcatcgtttggggcaacggcaccgggt ggcggtccgctggtttaccgtgatctgatgctgactggcggtcccagggtttgttggcgccggttaacacccc gatgagcaacctgggtatcagcatgcgtgctggcattgatatgctgaccggccgaacgctgatccggattccgtg ttctttgatcgcgttgacctggaacacattgggtgcgtccgggtcacagccaggcggtgccgcagcgattgtcgttgg ttctgaccgcgcatgacaccgtgctgccatccagccggcccgctggcgaaacatcgacgggttcgtgttcc cgactgctgctggcaggtcagcgcgactctatcgtttcccgcgctggtgaaagcgttctacaacgcccgcat cacattccggccctgtacggcgaaagtgcgtggcgctgaccattcaccgttgtaggtgatccaggccggttctgtg ctccgaccaccgctggttccgctcacctgatggcgatcgtaccggcgcgcgagttctcgttccgggct gcggtatttgcaccgataccgcgacctggagcgacgttcgtcgtaacagcctggctctgaacgtgccggcagcga ctccgtaagcggccgc
APET-4	catatggttggcccgaaaaagttcataaaaccgtgttctgccgagcagcgtgaaaggcttcggcgccggtacc atcttctaccggaccaacacccagggtaaaaccgtggcgcggttgttctgactccgggttcacctctaaaaaag atgatctgagctggtacggccggcgctggcggttccgtggcttcatcgttttcggtattgacacctggacccggcg gacttcccggtatgcgcgaccaccgaactggtggcgcggttgattttctgaagaaagataaggagctggcgca gcgtctggacccgacctgtgtccgttgcgggtactccatgggtggcgcgccgctgaaagcgcgacc aatgaccgggaatttctgagtgctggcgatcgcgccgttctacatcaaatgatcgggtctagcaccagcctgc cgttggtagcatcaccagcccacccgtgatcgttaccggcgaaaaagatgacgttgcgcggcggaactgttcg gccgtagcttctacgatggactggctaaaggcaccgcgctcagtagctgaaaggcgcggtatcacgata

APET-5

ccccggaacacaaaaacaccgatctctgagcacctctgttgcttctgaaaaccttctgtgataaagatgaacgt
tacgcgaaattcatcttcccggcgccgcagcgtccctgaccaccagcctgagcgcgcgttaagcggccgc
catatggcgggcgccgcgattacggcttgcgggcagctaccgggtgcttactctgatgttaccgggttgccg
gcggctgcagcgtgccggcgctgttactaccggcgaccacctctggcaccaaccagccggttgccgaccgg
tcgtttcccggtgatcgcttggccatggcttctccaggcgaccagcaaatacaccagcaccagcaccacctg
gcgagctggggttctgtgtgtgttccaaccacccagggtggcctgagcccgtcccacgggtgcgttcgctgacg
attgaaacaccggcctgacatggctggttgcgcaggtatggcaccaccggctcccgttccgtgatcacattaccac
caccgcgtggcggttctgttcacagcatgggagggcggtgcggcggtgttggcagcagcgagaaatccgaaa
attaaagctgtatctacttatgccgcagctgaaaccaaccgctctgctaaagcagcagcggtactctggtgctcc
ggcccagttctggcgggctctcaggacaccatcacccgggttggcgaccaccagcgtcaatgttcaccgcca
aaccgccggcacgtcagctgcgtaccattaccgggtggttccactgtggctttatggatagcagcggtctgttcg
gacaacgggtctatcaccgctgcggcgagctgaccgcgagcaaacgtgtgatgaccagctggtccgctacta
cctgtccggcgataccgcggcgctgacctggtgtggggccagccggcgcgctaccgatccggcggtgtgttcg
aaggcacgaataagcgccgc

APET-6

catatggcgaccaccgtttctacccggaaggcgccggcggttcccgtggttctgttcagccacggcttggtg
gccgtccggaagattacgcgccgtgctgcgtcagtgggcgccggcggttctgttgcggcgccggctttcc
cgacaccagcggcgccgcggcggttaccgacgttctggacgttatcaaccagccggcagatgcgtcttacgtta
tcggcaaaagtctggcactggatggaagcgtggcgatggcctgttccaccgcctggcgaccgatcgtgttcggg
ctgctggccacagcgcggcgcccatcaccaccgttggcctgttcacctggcgcggtgacgcgcgtctggatgcg
ggcatcgttatgtccggcagcagcctgggtactggcaccggcttccgtggcgccggcggtccgatgctgttcgtg
cacggccagcgtgatgaaaccgttagctacgcgagcggtaaagcggcatacgtatcggttccgtggccgaaag
ccatgctgaccttcccgaagcgaccacggcgcgacctgctgcgtgatggcgccgcgagtttcgaagtctg
ggcgacaccaccaccgaattcctgcgcttcgcgctgtacggcgatccggcgccgaaaccacgcacgtgaccgttc
ggcggtcgcacagatgtggcgacctggatgctgtgtgaagcgccgc

APET-7

catatgcccttcagcgtggtcctgaaccgaccaccggtagtattgaagcagaaaaaggtccgtttgccgtaactc
aggcaccgggtccgccaggtgacgggttcagcggcgccggaccttaacttaccggaatgacaccagccagggtacc
ttcggtgcgggtggcgatctctccgggttcatcagcccgaattcctgatcagctggtacggcccgctctggcgt
cccagggtctgtgttctgaccttccaccaacaccccggttgacacccgtctgcgcgtgcggatcagctgct
ggcgggcgtgaaatacctggtggaaaaaagccggctaaagaccgtatcgatccgagccgtctggcggtgatgg
gtcacagcatggcggtggtgtaccctggaagcgccggcgccgtaaccggcgctgcgtgctgcgatcccgct
gcacccgttccatcgcggcaggatttctctaaagtgcaccccgacctggtgatcgggtggtcaggttgatttctg
tgcgccggcgccgctgttcgctgttccgttctacgattccctgaaagcgaccagcaacaaaggctacattgaaatc
gcgggggggagccactcgtatgcgggtatcccgaacgtgaccatcgctaaatacagcatctctggctgaaacgc
ttcgttgataacgataccggttacgaacagttcctgtgcccgtgccgaaagatccgaaaaccagcgcgttcaag
gtaactgcccgcacacctaagcgccgc

APET-8

catatggggcccgccaccgactatcgctagcctgtccggtgatggtgtgttgaagtggcacctacaccgactccc
ggatgctccagaatttggcgatgccaccatctactaccgctgtccactaccatgccggttggtggcggttgcgattg
cgccgggtacactgaaaccagcgtcacatcaactggtggggcccgctctggctagccacgggttaccgtgtg
ctggttctggataccaacgaccgtcgtgaacgtcctgacgttcgcgcagatgcgctgatcggcggttcgtatcct
gaaagcggaaccaccggtgatggttcccactgcgtggtcgcattgatgatggcaaatggctatcatgggcca
ctctatgggtggtggcgccacgtgattgctgcgcgtgaacactctgacgaaattcaggcgccattccggttacc
ccgtggcagccggacggcgatttcgacggtatcaccgttccgacctcgtgatggcgggcagcgccgatcgct
tgctgacgcggcagaccatgcctggccgacttctgtctatcccgaagcaccaccaaaagttatatggaatc
gacggcggttctactacatcgccgacttctcgcggtccggatctggcgaccattggtcgttattaagcgccgc

APET-9

catatggctccgaccattgctagcctgtccggtgacggcggttttgaagtgggcacttacactgacttcccggatgc
gccagaattcggcgatgcaaccatttattatccgctgtctaccaccatgccggttggtggtgttcgattgcaccgg
gttacaccgaaaccagcggcacatcaactgggtggggcccgctctggcggtcccatggttacgcagtgctggttct
ggataccaacgatcgtcgcgaacgtccggacgtgctgctggatgcgctgatcggcgagttcgtatcctgaaag
ccgaaaaccaccgtgatggcgggcccgctgctggtggcgatcgtatggtgtaaaatggcgatcatgggtcactcca
tggggcgcggtggcacctgatcggcggcacgaacacagcgatgaaatccaggcgggcgatcccgttcacccc
gtggcaaccagacggcggtttcgacggcatcaccgttccgacctggttatggcgggtagcgcggtatcgtgttc
ggacgcagcggatcacgcgtggcgacacttctgagcattccgaaagcaccaccaaaagtatatggaaatcga
tgggtgttcccactacatcgggatagcagccgtggcccgatctggcgaccatcgccggttacggcatcgcgtg
gctgaaactgtacctggatggagatgagcgttaccgtggcttcatctacggcgaagaagctgaagccgatcgcga
caaatctctcgtatgttaccagcccgaagcgggcgc

APET-10

catatgagcccggcgccgaccgttgcgagcctgagcggcgatggtgttacgaagtggcacctacaccga
attcccggatgcggcggaattcggcgaagccaccatctactaccgatggatgctccgaccccgttcggcggtgtt
gcgatcgcggcggtacaccgaagttcagcgtcacatcaactgggtgggtccgctctggcatctcagggctac
gcggttctggttatgaacaccaacgatcgtatgctgatatgccggacgatcgtgcggttcgctgatcgcggcg
ttggcggttctgctgcggaacgatcggcggttagcccgctgttcggtcgtatcgtatcgggcgaatggcaat
catgggtcacagccgtggcggtggcgccacctgatcgggtaaacgaacaccgaacgaatccaggcgggc
cattccgttaactcttgggaaccggatggcgcggttcggaacatcaccgcggcaccctggttatcgaggcg
gaccgatcgcgcgtgttgcggaacacgcttggcgacacttctgtctatcccgaatctaccccgaagtg
tatatggaatcgggatgaaggctcacttcaccgctgatagcaccgtggcatggatctggcgatggtggcggtt
atggcatagcgtggttaaaactgtatctagacgggtgatgaacgttatagaggctttatctacgggtgatgaagctgagt
ctgcgcgtgataaattctctgttaccgttaccagcccataagcgggcgc

APET-11

catatgaaccgctaccagaaaggcccgatccgaccacgcgagcatcgaagcggacggtccgtacgcgggtga
gcaccagcgctgaaaggctctggttcggtagcggcaccgtttacagcccgaacgctgcaggtagctatgcg
ctggttgcggttgcggcaggcttcttccgcgcagagctctatcgcgcagatcagccgtcgtctggcgaccacg
gcttcgttgtggttacaattgataccaacacattttcgacttcccggattctcgtagcaccagctgctggctgcgct
gaaagcggttgcggcgctgagcaccggcccgggcgggcgaatcgaacggcgctcaggcggttagcgg
gctggagcatggcgggcggtggcacctgctcgcggcggttagcacgcccggcctgaaggcgcggttagcat
acgcgccgtgggcccgaacaacaacctgaaggcgtctaaggtgccggcgccatcctggcggtactgccga
catcgttgcgcctccggcatcttcgccgatgtgttctacaagcgatccggcggttccagcaaactgctgggt
atcgtgaaaggcgcggtacacttcttctaacactgcaagccagccggcgagctacaccaccgtagcctggttca
aacgttttgtgataacgacactcgtacggcccgttctgaacggcgacaccggttacagccggttcagctacacc
gtccggtttaagcgggcgc

APET-12

catatgggtccgaaccgacattctctcgttcgttctacggaccgtttgcggtttccagccagaccatttccggt
ggtaccttcttcggtggtgcgaccgtttacagcccgaatactccgggtcgttatgcgatcgttgcgttctgtccggc
ttaccgccaccgctcctcatcgtatgcaatggcgcgctcgttggcaagccacggtttcatcgtcgcagcgatgg
atacgcgtagcaccctcgtacgcaagcagccgcgtaccagctgatcggcgactgaatgctgttgcggta
tcaccaccggcccggtgcaggtaaaatcgaacggcccgccgtattgtgaccggtcactcgtatggcgggcgg
cggtaccctgattgctgctaccaacaaccttccgcaatcgcggcggaattgcgtggcgccgtggaacaccag
cgaaaacttcttggcaaccggttccgtccgctatcatcgttggcgcgctggataaccatgcaccggttgccttc
actccgtaccgttctacaacgatgaccgtaacgaactgctgtgctgatcagcgcgagccacttctt
cccgcaggatctccgccgaaccagccggcgtctgcgtgcagatctcctgggcgaacgttccggcatggcg
ataccggttacggcttcatcctgaacgcacgtccggcggaatgtggcagctggcgaccaccggtccgttctaag
cgggcgc

APET-13

catatggcgggtacagaacgggcccggcaccgacgctggacgcaatcaaagctgatggcccgtatcgatcacctc
tcaggcagtggcaggtaacggctttggcgagggaccatctacaccccaccgcccgggtaatacgcactgg
ttgcgatttccccggcttcgttccgcgagagcaccatggcagggctggcgaacgtctggccaccacggttt
tgtgtgtgtgtatcagcaccaacacgctgttcgacttccgtccagcgtgagaccagctgctcgagcgtg
aaaaccgttagcgccttaaacaccgggtgccgtggcgggcaaatcgacaccaccgccaggcgggttctggctg
gtccatgggcccggcggtaccctggaagcggcaggttctacgccgggtctggtggctgctgttgcgttctcc
gtgggataccagcaccacaaaatgaaagttaacaaagtccggcggcgatcatcgggcgacctccgacacc
atcgcgccgggtggccagccactcccagaaattctacgacgcgatccggcgaccaccccgaactgctggcagt
tatcgaaggtggcgatcacttcttcccgacaaaagcgtctgaaccgaccagctacaccaacatctcgtgatgaa
cgcttcggcgacggcgataccggttacgacaccttctgaacggtaacgatcgggcgctggcgagcttcagcag
caacgggtccgttctaagcgccgc

APET-14

catatggcgttcgcgatgaaatggccaggcggcaccgcccgaacatcaccgggtgatggcagcttctctac
cagcagcagcgcgatcttcggccagagcgggttcggcgggcgacagcttactacccgaacaccgcccgttaa
acccgggtgttccttcgcggcggttctgagcgactggaacgcgctgaactggctggcccgcggttgcga
gctggggcttcgtgtgtgtgtgttaacaccaacaccccgttcgattccccgatgctcgtggcgatcagctgctgg
cggcgctcaactggcggttaacagcgccggcgagcgggttcgtgataaagctgatggctcccgctggtgtta
ggcggtggagcatggcgggcggtaccctggaagcactggcgaaagataaccaccgggtaccgtgaaagcgg
gtgttccgctggcgccgtggcacaccaacaaacctggagcaaatgacgaaccgggtttcatcatcggtggcc
agaacgattccgttcggcgccggcgagccacgctatccgttctacaacagcctggcggttcggaattctacc
tggaaactgctggcgcgatcacttcttcccgacaaaagcaacgggtaccgttagccgtcggggttttattctca
aacgccacgttctgccgataccggttactccgttctgtgtgttttctggctcggatgttagcaacttctgtta
tcttgctaagcgccgc

APET-15

catatggcgggcttcgaaaaaggcccggatccgaccaccgctctgttaccgcgaccgcccagctacgatat
cgcgaggttgatgttccgaaaggcagcgccaccggcttcgcgatggcaccatctactacccgaatctactgc
ggatggcactttcggcgcggtggcggtgattccgggcttctggaaccgaaaagcgcgatcattggtacggctc
gcgtctggcgagccagggtttgtagtgtgacccgtggacaccaccagcgtgttcgacactccgtagaccgcg
gcgggtcagctgctggccgcgctggattacctgacttctaaatctaccgtgaaagatcgtatcgatccgaaccgctg
gcgggtggtggccacagcatggcgcggtggcggtaccctgtccgcgcgccgaaccgtccgaacctgaaagcg
gctgtggctctggcaccgtggcacgtgtccgtccgtgaccagcgtcaaaaccccaccatgatcttcggcagc
gacaacgactttatcgaccggacggtagcttccggaacccgtttacgatggcatcagcaacgcgcccggagaa
agcatacgttctgtaaaacgcgggccacatggcttacgttttccggacaaacgtatcgcggaatacacctg
agctggctgaaacgctacgttgatgatgacaccggtacaccagttcctgtgtccggggccgaccaacgaaagc
accatcgggcggtataatccacctgcccgatgtaagcgccgc

APET-16

catatgccgtctgcaccgtgcagcaactgtaccgtgggtccagctccgactgctaacgctctgaaagctagcaccg
gtccgttcagcaccgcgcgttctctgttagcggttctgaaaggcttcgggtgatagcaccgttactaccgacca
acaccaccggcaagatggcgcgatcgcggtatcccaggctacctgtcttacgaaagctccatcgaatggtggg
gcccgcgtctggcgagccacgggttcgtgttatgacctgaacaccgaaacctctacgatcagccggacagcc
gtgcgagccagctgtccgaagcgtggattacctgatcgcgcagtctaaaagcagctccttccgctgtacaacaa
aattgattctaccgctggcgcaatcggttggtctatggcgggcgggcagcccttaaactgagcaccagcg
tagcatcaacgcgatcatcccgcaggcgccgtggtacttcggctctaactccttcaacaccatcaaaacccggcg
ctgactctggcgtgctctcggtatccgttgcggcgacgcgagcccgttctacaaccgtatcccgg
cgagcaccggaaagcgttctggaaatctacggcgccagccacttctgcggaacagcggctaccggaacgaa
gatctgctgggtatgtacggcatcgcttgatgaaacgtttcatcgatttcgacagccgttacagccagttcctgtgc
ggcccgaaaccgaaggcgatctggcgatctctgaataccgtgaaactgcccgtactaagcgccgc

APET-17

catatgaccaatccgtatgcgcgtgggtccgaacccgactgcggcgctactggaagcgagcgcggggccgtttac
cgttcgctccttaccggttagccgtccgtctgggtacgggtgccggtaccgtttactaccgactaacgcagcgcgta
ccgttggcgcaatcgccatcgtgccaggctacaccgcccgcagagctctattaactggtggggcccgctctgg
cgccccacggcttcgtagtattaccatcgacaccaattctaccctggatcagccgtcctctcgtagtagccagcag
atggcgggccctgcgtcaggtggcaagcctgaacggcacctctagcagcccgatctacggcaaagttgataccgc
ccgtatggcggtgatgggctggagcatgggtggcggggttctctgattagcgccgctaacaacccgtccctgaa
agctgctgcggcgaggcgccgtggaacagctctaccaacttcagcagcgtgaccgttccgacctgatattcgc
gtgcgaaaacgactctatcgccgggttaacagccacgcgctgccgttctacaactccatgagccgtaacgcgaa
acagttcttgaaatcaaccgtggcagccacagctgcgctaacagcggcaacagcaaccaggcgctgatccgt
aaaaaaggcggtggcggtgatgaaacgctcatggataacgatacccggttacagcaccttcgcgtgccagaacccg
aactctaccggtgttctgatttccgtaccgcgaactgcagctaagcgggccgc

APET-18

catatgtctacctgtaccgtaaacgctggtattgaaatcgaccggttccgtccctgtctttgtgcgtgtagccagg
gcccgtgcgggttagcaccagcaccgtgagctcttctgttgatggcttcggcgggcgaccatctaccacccga
gcaacatctctgatgaaatggcggtgatcgatcagcccaggcttaccgcggcgagagctctatcgctgggt
ggggcccgctgctggcgagccacggttctgttgatcaccatcgataccaacggccgcttcgatcagccacca
gcccgtggccgtcagctggatagcgcgctgtcttacctgatcagcgaaggtaacgtagcaactctccggtgagcg
gcctgattgatgaaaaccgtctggcgaccatgggccacagcatgggtgggtggcgggcgctgcagagcgccgc
gcgtaaccgtctgagcgcgtctgttccgtggcgcccttcaacaccggcagcaacgagttcaaccagatcaacgtt
ccgacctgatcatgggtgcgaaaacgataattgtggcgccggttaactctcacgcgaacccgttctacggctgat
tccgggaagcaccgataaagcgttcttgaaatcaacaacggtagccactcctgcgcaaacgggtgtagtcca
acaacgcaaccctgagcacctacggcggtgagctggatgaaacgttcttgatcaggatacccggttacaaccagtt
cctgtgcgggtccgaaccacgttgcggatagccgtatcagcgatttccgtgatacctgcgcgttctaagcgggccgc
catatgctggacagcccgtaccagcgtgggtccggacccgaccgaagcgagcggttaaacgagatcgggggccgt
tcggggtgatcaggtgaaagcgaccggccacaccggttccggcagcgccaccatcttctaccgaaaatgaaa
aaggcaccttggcactatcgccgtgagccctggcttcatttctccggaaggcctggttactggtacgcgaaaacc
ctggcttctacgggttttgggtgatggcgatgggtacgttaccgtagcggatccggccgagccgtggcgatc
agctgcgtgcagccctggattacatgaaaacaaaagcccggttaagataaagtggatcccaccgctctggggc
tgtccggtcactctatggcggtggcggtaccctggatgcggcgctctaaaatgccggacctgaaagcggcagttc
cgctggcgccggtggtatctgaccggcggtccggttatccgaaaagttaagcgccgacctgtggttggcgcg
aacatcgatttctgtgcgcgatggcgatcttcgcgacgcggttctacaagcgctgaccttccgaacaaagcgc
tgctgacctgcccgggtgatcacatgaccccggttctgtggcgaaaccgacctgaccaaattcgcggtcagct
ggttcaaacgttcttgatgacgatacccgctacgatcagttcctgtgcccgaaccggcgcttctacgaaggtacc
tgcccgtgacctaaagcgggccgc

APET-19

catatggttgcggcgcggttaccaaaggcccggtacgctctcggtggaagttgctaagtaaccgtagctac
gccgtgaccgaaaccaaagttgacgatcgggcgcttaccaaattcggtccggcgaccgtttacgcgccgaaagg
cgctccgggtgaaaccttcggcgcgatcgcgatctcgccgggttcttggtgaacagtggacctgacctggct
ggcacagcgtctgagctccttcggttctgtaccatcgttttaccgccgaacagcatcttcgaccagccggtatgcg
gtagccgtgcgctgctggatgcgctggaaacctgaaaacccagcttaccgcgaaagcgctgatcagccgagc
cgctctggcggtgctgggccacagcatggcgggcggtggcgctggaagcggcgagccgtgatctgacctg
aaagcgggcggttgcgctgctgccgtgggacctgactaaaaactttccggcgcttcgacccccgacctggttatcg
gtgcaaaacgggatttcatcgcgccggttgcgagccacgcgaaaccgttctacgctagcctgagcacctctctgcc
gcgtgcgtacctggaactgaacattgggtcacgcgcaaccggccctggttccgaccaccgaagtgcccgtgcgg
cagtgaaactggttcaaacgtttctggacgacgatcctgttacacaaaactctgtgcccgccggtggccgggtt
aacgcctctaccgtaactgcgtatcagtcgtcctgtggtcagggcccgctgtcttaagcgggccgc

APET-20

APET-21

catatgaacccgacgctgggtgctctgatcgccggtgcgcaaccgcttctccggctgttggtcagggctccggcg
caaccatcaccagcctgagcgggtgacggcggttacgaagtgggcacatataccgacttcccggatgtgccggaat
ttggcgaagcgactatctactaccgatggacgcacctggcccgtctggcggcggttcgatcgccgggttaca
ccgaaatccagcgtcacatcaactgggtgggtccgcgtctggcttctcacggctacgcccgttctggtatgaacacc
aacgatcgtatggaagatatccggcgggaccgtgcgggtggcgctgatcgccgcggttcgtctgctgcgcgga
aaacgatcgtcgtgattccccgctgttcggccgatcgtatgctgctaaaatggcgatcatgggcccactcccgtggc
ggtggcgccgaccctgattgcagccaacgaacacccggagaaatccaggcggctatcccgttaacagctggg
aaccggatggtgtgttcggaacatcaccgcgcgaccctggtgatcgccggtgcgaccgatcgtatcgcgct
gttgcggaacacgcgtggccgacttctgagcatccacgaaagcaccggaaatgttacatgaaatcgccgat
gaaggtcacttaccgcggacagcaccctgggtatggatctggcgatggttggccgttacggtatcgctgaagcg
gccgc

APET-22

catatgtacgacccgatggtgcacttcatcctgttggtgtactgtctcgcactgtttgacaatcccgtgatcgtcc
attgcctactgaagtatggtatccggctgttcagaagcagcagatacgccgggtgagcgtgttggcgctcggct
caccggcgcatcgacgcgacgccggcggaggcgcccgatccgctgatcgtgttctcccacgaaacca
gggtattcgtgttcagagcaccttctacaccgttcactggcaagccgcggttacgtgtgttgatggtatcaca
ccggtaacaccttcagcgacaacagccaggcggtgcgatcgccgcgtaaccgtccggacgacttctcttc
gttgcgctaccgtgatcgtatcgctgcgtgagcggctccgctggcaggcatgatcgtatcgtatggtg
cagttaccgggtcacagcttcggcgcttgacccctggcggtgtggcgcggtattctgttccgtgcggcgct
gccgctggcggtgccgcacgacccggcgatcatcgctccggtttctatgatcgtatcgtatccggttatgctga
tcggcgcgagcgtgaccagaccacccgtacagcatgtctgaagcgggttacgatcgtgcgaccccgcgctc
gtctgttcggcatcgataacgcgggtcacttcttaccgacatctgcgcgtgggtgcgattccggcgcgga
cgaagatggttgcggtgacgattccgtccggcgagcgggtcaggcgctgaccaaccgtctgtctagcctgttc
ttcgatcagcttctgaaacagctgccggaagcgcggtgcgctgcaggcggtatcgtgttgcggcggttcgcgga
cccgtggtatgtgtacctgagcgattaagcgccgc

APET-23

catatgaccctgaatttaaaccgcagccgggcccgtgagcatcaaaagcgagcggcgatcagcttccgtgtt
gaagatcaggcgaaagatctgcagatccgcgttacctacccggatggtccgggcccgtggccggcgatccttc
tctcaggtgcgatgtccagcaaagatttctataaccgtgttgcggatcactggcgacccacggctatgcgacat
cctgccgacccacctggatagcgaagcctgggctacttctcaaaaaccgctgccgcgtgaaaagtttcttcg
gccgtatcggcgatagcagttcattctgaaccacctggacgaaatcgcggttaaagctggcatcggggcaaaa
tcgaaaaaaccaaactggcagttgctggccactccttcgggtggctggaccagcctgatcctggctggcctgccggt
taaaatgccggagcggcagcctgaaaagcttcgtgactctctgatcaaaagcgtggtgagctataacggcattggc
ccgatggatgaaatcccggatgcagcatggcgaaaagttgcaatcccgggttttcgcgcgagcggcgaccaacga
tccgggtgcgacccgctctggcccgggtcacccgtggcggttgctatcggtgcgtacgactacgcccgttccaa
agaaaaatacgtgtgagcatcactatgggtgaccactactacgggtggcctgatctgccgtgaaggcgcgggcg
gcaaaccggacccggaaggctgtgtccttcgttaacggtagcagcaccgcgtcatggatgcttacctgaaagatga
taaactctgtaaaaaattcctgcgtaccgcagacctgcgtccgctgacgaatgaccgtcggttccctgaacgtagct
aagcggccgc

APET-24

catatggcgccgggtgattggagcaaaaccggtccgcaccagggttaaagtgataaagcggacgcgcacaccatc
taccgtccgtctgacctgggcagcggcggtgaaaaaccgggtgttatctggggcaacggcaccggcgctac
cccggcgcgctacgatggcctgctgcgtcactggcgagccacggcttcgtggttgcggcggaacacaaa
caggcgaaacggtaacgcgatgcgtgcgggtctggtatggttacaaatggaacaccaccagctccagccc
gtactacggcaaaagttgacgttagccgtgttgcgagcagcgtcactctcaggcgcggtgcgggtgcgatgaacg
cggcagtggtatccggtgtggacgcggtgtccgatccagccgggcccgtggcgcgccgcgtaaaftcaa
acagccgaccttctacatggctggtcagaacgatcgtgtgttggccgagctgggtgcgtaacttctacaacaaat

ctggccacctgccggcggtgttcggcggtgttcgtggcgacccacgttgcctggcggtgatggtggagcgt
ttagaggccaacaaccgctggtgttccacctgcgcggcgatgcgtcgcgcggtgtcagttccggcgcg
gtcctgtggctactgcgggtgttctccccgttccaactggctgcgaatactgtttaccagccagaacggttaa
cgggccgc

catatgaccgcggcagcggtttccagccgctactacgaagagaccgggtccgcacgcggtgaccgttgaacagg
gtggtccggatcatacctgtactatccgagcgatctggcgcgggcgccacccagcaccgggtgtgatctggg
gtaacgggtactggcgcaaccccggaacctatgatgaaggcctgcgtcacttcgcgtcctggggcttctgtggtgc
agctgccaacaccactcaggcagcgagcggtaaagaaatgctggcaggtgccgcttctgattgaagaagatc
agcgtagcggctgttttctacggcaaaatcgacgaaactcgtattggtgcggcggggtcactctcaggggggag
gcgggtgtatcgtgctggcgcggtatccgattgtcgataccaccatcccgatccagccggggcccgagggcact
gtaggggacctgcacgggtccggcggttcttctccgggtcagttcgactggatcgtccgctgagctacgtttacag
ccgttatctggcgccgatcacgttccggcagtggtcgggaactgcgtggcgctgaccacttctccctggcgaca
cccgctgcgcctgatcgggtgcgaccacggcggttccggttccacctgatggcgagcaaaacgctcgtggtat
cttctcgcccgctcctccactgcggcatctgcactgacaccgaagcatggagcaaaagtggtatgtaacgcgaa
agcgcagggcattccgggttaagcggccgc

catatggatccgtgggatgcgcggggcccgtagcgaccgcggttcagatcgggtgcggttaacactctgtattacc
cgcgtagatcgcgtctagcccgcgccgtcaccgggtatcgtgtgggtaacggcaccgggtgtgcgcgcgtgg
ttaccgcgatctgctgctgtactggcgagccacggatttattgtcagcgagcaaacacggcggttctaactt
ggcatcfaatgcgcctcggcatcgtactggcagcagccgacgcggatccgggttcacattccgtggcaaa
gttgacctggcgacatcgggtgcgagcgccactctcagggcgcgcgggcgcgatcgttgggtgcggacc
cgcgtagcgtgcatcctcggcatccagccgggtccgctggcgacatcgtatgttcacgttccggcgctgct
gttggcgggccagcagattccatcgtttccggtcctgtttaaagcgttctacaacgatgcgaccacatcccgg
cgatctacgcggaagttcgtggcgcggtacattcaccgttgggtggcgatccggggccggttcggcgccgacca
ccgatggttccggttccagctgatggcgacaccggcgcggtgcggaatttatggctcttctgtggtatttgca
ccgatactcgtacctggtctgatgttcgtgtaatgctctggttccgtaagcggccgc

catatgggtaccggtttccgggtgtttctaattctcgtcgttggcagcttctgctaccaccagtgggtccgagcg
catcttgacagttttcgcccaagcaccttggggcgaacaaccgtaaacaccgatcatcgttggggcaacgg
tactactgcctctccgaacacctacgccagcctgcttcagcactgggcttctcacggttctgttctgcggcgaa
cacctctaaccggggcaccggccaggaatgctcggctgcgttgattacctgaccaccagaacaaccgttctac
cggtacctacgctggtaaactggatctgaaccgtatcgcgcgagcgggccacagccagggtggcggtggcacc
atcatggcgggccaggattaccgtatcaaggttaccgcgccgttccagccgtacaccatcggtctggggccacaac
tctagctcccagagcaaccagaacggcccgatgttctgataccggtagcatcgataccatcgcgagcccgac
gctgaacgcgctgccggtttacaccgtgcgaacgttccggtttctgggggtgaactgtctggtgtagccattcg
aaccggttggtaacgcgggtgatttccgtggccggcgaccgcgtggttccgctaccacgtgatggatgatcgga
gcgcggaagataccttctacggtagcaactgcgatctgtgcaccgatcgtgattgggacgttcgtgtaagggtat
caactaagcggccgc

catatgggcggtggcgggcgaagataacgtttgtgaaccgggtgatttcgacgttgtgaccgagccggcggt
cgcgattgcaccatctccgctccggagcagctgggtgacgaccacgccgtggttatctggggcaacggcaccgg
cgcgagcccgaacacttacccggcctgtgcgtcactatgctcccacgggtttgttgttgcgtgtaacacca
gcaacgcgggactggtgaagatatgctgaactgcctggattttctggaaggctcactggttcttccgtggac
ttctctaaagtgggtgaaccggctcactctcagggcggtggtggtgcgattatggcggtgttgcgcgcacgc
tactaccgctccgattgaaggcttcacctgggtctgggccacgatcgtgattcccacgaccagcagaacggtcc
gatgctgatcctgtccggtagcgcgaccgtctggtgaaccgcgtaccaaccacgcgccgctgtttgatgcgac
caatgttccggcggttctggggcgatcgcgagggtgccggtcatttgaaccggttggcgatggcgcgacttccgc

APET-29

ggatctccaccgcctggttcctgtccagctcgaagatgacccggaagcgaccacccacttcgtgggcgatgatt
gtttcgtttgcgaagatccagggttcgatgtgaaggtgtgaaagaaatcgattaagcgccgc
catatgcaggataacagcgatggtcgttacggtccggaaggccgtaccttcgcgcgtggtgcggcgaaactgttc
gatgatcgtctgttcgaagcggtacactacgttacgggtaccgatgttcggaattcgccagcgactatctctac
ccgctgacctgtcttcgatctgccgaacggcggtgtgttatgagcccaggctacaccggcaccgccgaatccta
cgattggtggggtccgatgctggcgctgtgtggtatcatcaccgctatcatcgaaaccaacaccccggaagataac
ctggaacagcgtaaaacgcgctgatcgcggcatcgaactgatcaaacgtgaaacgcgaactctgcgagccc
gatcaacaacaaagttgacgaaagccagatcgcgatcatgggtcacagcctgggcggcgccgcgagcctgcgt
gcagcggaagaactggaaggcgacatcaaagcggttatcccgctgaccccgctactgttcgaactgggccaga
gcttcgaaggcgacctgaaggcggtgagcaccgccacccgtgatcattgcggcggtgacgacaccatcgacc
gccggaaacccacagccgtctgctgtacgacgaaatcaactcctccaccaacaaagtgtacctggagtttgcgga
aggtgcgcacggctgccgaccaaccaggggagcgaattccagaccagggcacctacgtttacgcgttcctga
aagccaactccaccaacgacagcaagtacagcgatttcacggcggtgacggcgaggactccttcagcagctatg
aaactaactaagcgccgc

APET-30

catatgccggggccgcacctgactccgttccgtattaccagctccgacacccgtggctgcagaaacggcagccg
ctggacctgcgcccccggtagcgggtccgggttcggtaacgggtaccatctactaccgactgacgactctca
gggcacttttggtgcggtaacgtgattccgggttcttgggtggtggaaggcaccctgcagtggctggggccgcg
tctggcgagccagggtttgtctgtgatgacctgaacccgttgacgccgctggatccgccgtttaccgtggtgatg
aactggttgcgtgcgtgaaatactgaccggcaactcctccggttaagaccgtatcgacccgcacgcctgggcgt
cattggctggtctatggcggtggcgccgactgcgtgcggcgccgggacatccggacatcaaagcggtatgtg
cgctgagcccggtgggatacgggtgggtgtggtggcaggttagcaccggacccgtgttttcggcgccgactggg
acgcaattgcggcaccggcggttttcggcggtccgttctacgatggcctgaaaacaaagacaaagcgctgatcc
agctgcgtaacgcgaccactccactacatggtgccgaacaccccgaccgcggaatacagcctggtgtggctg
aaacgtttcatcgacgatgataccggttatgataaattcctgtgtcctgctcgggtactggtgatgccattagtga
caccgttcagtgcgcgtgccgtaagcgccgc

APET-31

catatggttgttctgggtacgcgtgcgatggcggtaacaccagcacatcaccaggaagcggttagcatcagc
ggcgcgaccaacagccagctgtactaccggaccggtatctccaacgtgaccggtgcacctccatgtccgcggg
ctacaccagggcctggcgaacgttgatggctgagcaaaaaactggcggaacacggttatgtggtggttcggtt
caccgggataacacctgggctatgtttctcagtggcgtgacctgcacaaagcggcgattcagaaatcaaagat
ctgaacaacaaccacagcgtgctgaaaggtaaatcaactgcgaaactgcagacctgcggtcactctaaaggc
ggcgggcggtctctgtggcgagcgcgcagctgggttagcaccctgaaaaccaccattggcatggcgccgtacg
aagaaggcttcgctgatagcgcgtggcgagcatcaccgcggcgaccttcacaggcgggcggtggcgatac
cctggcgaccaactctatgaccggtggcgaatacagcggctgagcaccagcatctctctgtatgtacgttgaatac
accggttacagccacctggcgtgggtgagcgtgagcggcggaagcggcgagcgtgcgacctgtcgaaaccga
atctccggcgacatcatcgctggatgaaatattacctggacggccagacctcttatgcgagcaccctgttaacac
cagtggcaccgcgcgccacgagtggttgacaaaagcggttcttcaagctcctccagcagcagctccagctctag
cggctgcaactaagcgccgc

APET-32

catatgccaagcccgaaggccaggctctgggtgtaccggcgatttcgaaatctgctcttacgaagcgaacctg
gaaaacgatggttacgcgtctgctgatgacctaccggtgcgatctgtctggtggcccgattccggcgaccaccc
tgaccggcggtttaccaacaccaaagaacagatggaatggctggcggaacacctgagctctcacggctacgcg
gttttgacctgacccgaacaacacctgggtgtgcccgccgggtggcggtgatgcagattcgccggcttcgc
ggaattcgaagatgaaaacagccgttctggcagcccgtgcagggtaaaatcgatctgaccaaactgaacctgat
gggctttctatggcgggcggtggtgtatcctggctgcggaagaaatggttgaccagcacacctctgcgatcgcg
ctggcgccgtgggtgggtgcttactgttgatttcgaagcaatcgaagcccgatcctgatgctgggtctgaattc

gatgaactggcgtagccaccgaacgttactacagcgaactgagcccgaacaccgaacggcctggcgatgt
 acgcgggcggttctcacttccactgggtcgtaacagcgagaccgaatacaaaaagaattccgcaccctgggtga
 ccgattctctggaagttcagctgaaaggcgatacctccgcttacgacttctgacgggtgcggaacatgaagagc
 acgttgaagagggttgggtctctgctgacgattaccagaataagcggccgc
 APET-33 catatggcaccggcagcagcggcaggactggcccgtatgatgctgaatatgaaaccaccttccgtctgtacggtc
 ataccatctaccgtctgccaactgcccggacgatgagcaactgcctatcgttctgtggggcaacgggtggctgcat
 ggctgacggtagctatgttcgatgaattctgctggaatcgcaagccacggcttctgggtatcgtaacggctgccc
 gggcgggcagcggcagcaccgatgcagatatgttaacgggatgcgactggcggtgtagacgaaaacgatcgt
 ctgttcggtagttatcgtagccacctggacaccggagctatccggtgatgggtcagtcctgtggtggtattgaagc
 gtacgaagcatctgatgatgaccgtgtggataccaccgttctgtggaacagcggatgctgagcgatctggataact
 ggctgtgaaccgtctggaagcggcgatgcgctacttcaccggcggtccggatgacatcgcgacgaaaacgc
 ggttgatgattgggatcgatcccgtccgatctgccggcggtcatggcgagcctggatgttggtcacaccggcacct
 tcggcgaaaccgaacggcgggcgaatacggcgtgttgccggcggtggctgaaatggcagctgaccggcgatg
 gccaggcaggcgggcggttcaccggctctgattggcggtgtgtagcactgattgggatgttcgtcgtcgtaatct
 gtaagcggccgc
 catatggcttacctgccggcgaaacctcgaagctgacaccttccgggttacgttgcacgatccggcgcggtga
 ccgcgatatcccgatcgtggttcgttatccgcgcgatgctaccggctccggtggtggtttctctccgggtg
 gcgcgtacgttccgctgcgaccggctattctgaatggggcacctgctggcacgtgcgggttacttcaccgttaa
 cctgagccaccggaaagcggcgggcggtgcgactgtgcgcgctggcgatcccgcgagcgaatgcgatgc
 ggcgcactttgttatcggtaaccggcggaaggcggtaccctgaacggcccgtgggtaccaccgtccgatggatg
 APET-34 cgcgtgcgggttctggacgacctggcgggcgatgcagaagccgctggcggttctgttgatctggatcgtattgcgat
 cgttggtcacagcggcggggtttaccaccatggcgctggcgggcgcgcgatcgattactccccgagcgttc
 gtgcagttgacatgagcgacccgcgtccgcgcgcttccctggccaactccccgcaggcggttggtcatgtgggtc
 tgaccgaaaccagctgggatctgatcggcgcccggtgatgaccagaccgggtcgtacgatcgtaacgatcgtaagggc
 gaaccggcgagcggctcgtggatgcggcgatgcacatgccggcgccagacaaactgacctacctggaca
 gcgaagatgctactcactctgtttcggcctgggcccggacgcgtcgcgatgctggaactgtacgtgggtctgc
 aggtgtcgcgttttagatgcctatgtcgtgatagatctgatgctgcacatggctgggttctggtagcattgatgaa
 tggctgatgggtgtagctaccatcatgtctcgttaagcggccgc

Supplementary table 2. Data collection and refinement statistics.

	<i>Kb</i> PETase
Wavelength	0.9792
Resolution range	31.33 - 1.75 (1.813 - 1.75)
Space group	P 1 21 1
Unit cell	62.496 104.798 71.02 90 93.829 90
Total reflections	258810 (26166)

Unique reflections	89061 (8687)
Multiplicity	2.9 (3.0)
Completeness (%)	96.96 (94.87)
Mean I/sigma(I)	13.42 (3.54)
Wilson B-factor	17.88
R-merge	0.06559 (0.5443)
R-meas	0.08031 (0.6599)
R-pim	0.04582 (0.3703)
CC1/2	0.997 (0.649)
CC*	0.999 (0.887)
Reflections used in refinement	89046 (8687)
Reflections used for R-free	1019 (100)
R-work	0.1647 (0.2686)
R-free	0.1817 (0.3026)
CC (work)	0.970 (0.846)
CC (free)	0.969 (0.795)
Number of non-hydrogen atoms	8288
macromolecules	7465
solvent	823
Protein residues	1004
RMS (bonds)	0.014
RMS (angles)	1.60

Ramachandran favored (%)	97.59
Ramachandran allowed (%)	2.41
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.39
Clashscore	3.44
Average B-factor	20.84
macromolecules	19.75
solvent	30.69

Statistics for the highest-resolution shell are shown in parentheses.

References:

- 1 Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology* **37**, 420-423, doi:10.1038/s41587-019-0036-z (2019).
- 2 Furukawa, M., Kawakami, N., Tomizawa, A. & Miyamoto, K. Efficient Degradation of Poly(ethylene terephthalate) with *Thermobifida fusca* Cutinase Exhibiting Improved Catalytic Activity Generated using Mutagenesis and Additive-based Approaches. *Scientific Reports* **9**, 16038, doi:10.1038/s41598-019-52379-z (2019).