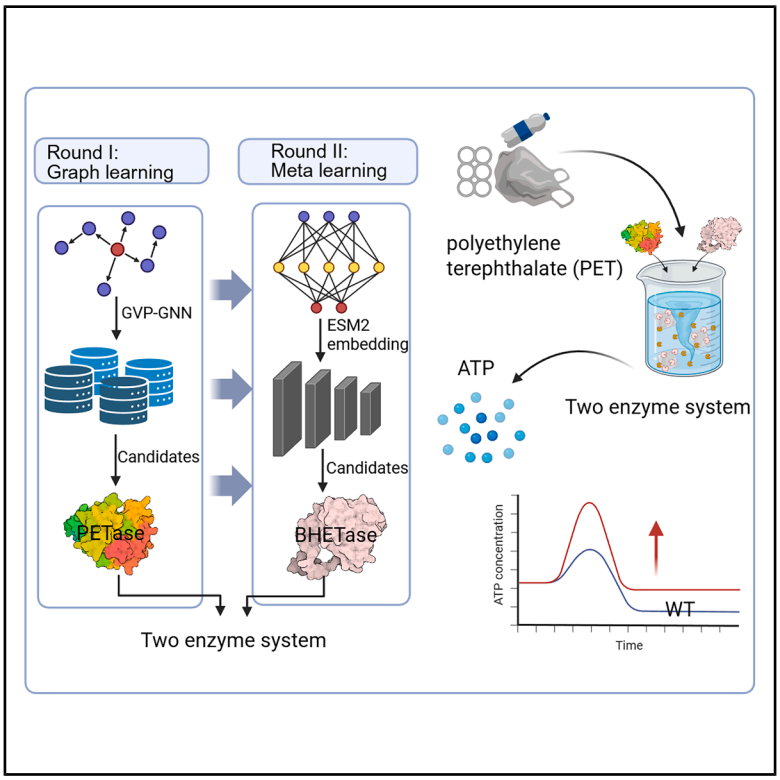
Cell Reports
Physical Science

Article

Integrating graph learning and meta-learning to enhance PET hydrolase activity at elevated temperatures

Graphical abstract



Highlights

- Two-round machine learning discovers thermostable PET hydrolase variants
- Meta-learning improves mutation compatibility and catalytic efficiency
- MD-driven two-enzyme system achieves 1.5-fold increase in TPA yield

Authors

Anni Li, Yibo Song, Jie Hu, ..., He Huang, Shengying Li, Xiujuan Li

Correspondence

lishengying@sdu.edu.cn (S.L.),
lixiujuan@njnu.edu.cn (X.L.)

In brief

Li et al. present a two-round machine-learning framework integrating GVP-GNN and meta-learning to evolve robust PET hydrolases. The engineered variants exhibit enhanced thermostability and catalytic efficiency, while a two-enzyme system achieves a 1.5-fold increase in terephthalic acid production from PET.

Li et al., 2026, Cell Reports Physical Science 7, 103037
January 21, 2026
https://doi.org/10.1016/j.xcrp.2025.103037

CellPress

Cell Reports
Physical Science

CellPress
OPEN ACCESS

Article

Integrating graph learning and meta-learning to enhance PET hydrolase activity at elevated temperatures

Anni Li,[1,5] Yibo Song,[1,5] Jie Hu,[3,5] Hejian Zhang,[1,2] Zhanwei Li,[3] Jin Tang,[3] He Huang,[1] Shengying Li,[4,*] and Xiujuan Li[1,6,*]
[1]State Key Laboratory of Microbial Technology, School of Food Science and Pharmaceutical Engineering, Nanjing Normal University, Nanjing 210097, People's Republic of China
[2]Ministry of Education Key Laboratory of NSLSCS, Jiangsu Key Laboratory for Pathogens and Ecosystems, College of Life Science, Nanjing Normal University, Nanjing 210097, People's Republic of China
[3]Research Center for Life Sciences Computing, Zhejiang Lab, Hangzhou 311121, People's Republic of China
[4]State Key Laboratory of Microbial Technology, Shandong University, Qingdao 266237, People's Republic of China
[5]These authors contributed equally
[6]Lead contact
*Correspondence: lishengying@sdu.edu.cn (S.L.), lixiujuan@njnu.edu.cn (X.L.)
https://doi.org/10.1016/j.xcrp.2025.103037

SUMMARY

The identification and design of efficient biocatalysts to address plastic waste accumulation are pressing global challenges. Leveraging recent advancements in artificial intelligence (AI), we developed an integrated machine-learning framework incorporating the geometric vector perceptron-graph neural network and learn2learn models to yield robust and thermostable variants based on three representative polyethylene terephthalate hydrolases (i.e., FAST-PETase, DepoPETase, and BHETase). The engineered hydrolase variants demonstrated up to a 5.5-fold improvement in hydrolytic activity across a temperature range of 40°C–70°C compared to the wild-type enzymes. Molecular dynamics simulations revealed key microenvironmental interactions sustaining hydrolytic activity at high temperatures—enhanced hydrogen bonding and hydrophobic interactions around active sites—and highlighted the unbound dynamics of bis(2-hydroxyethyl) terephthalate and mono-(2-hydroxyethyl) terephthalate near catalytic sites, motivating a two-enzyme system that achieved a 1.5-fold increase in terephthalic acid production. This study demonstrates a scalable AI-guided strategy for evolving robust polyethylene terephthalate hydrolases for industrial biodegradation.

INTRODUCTION

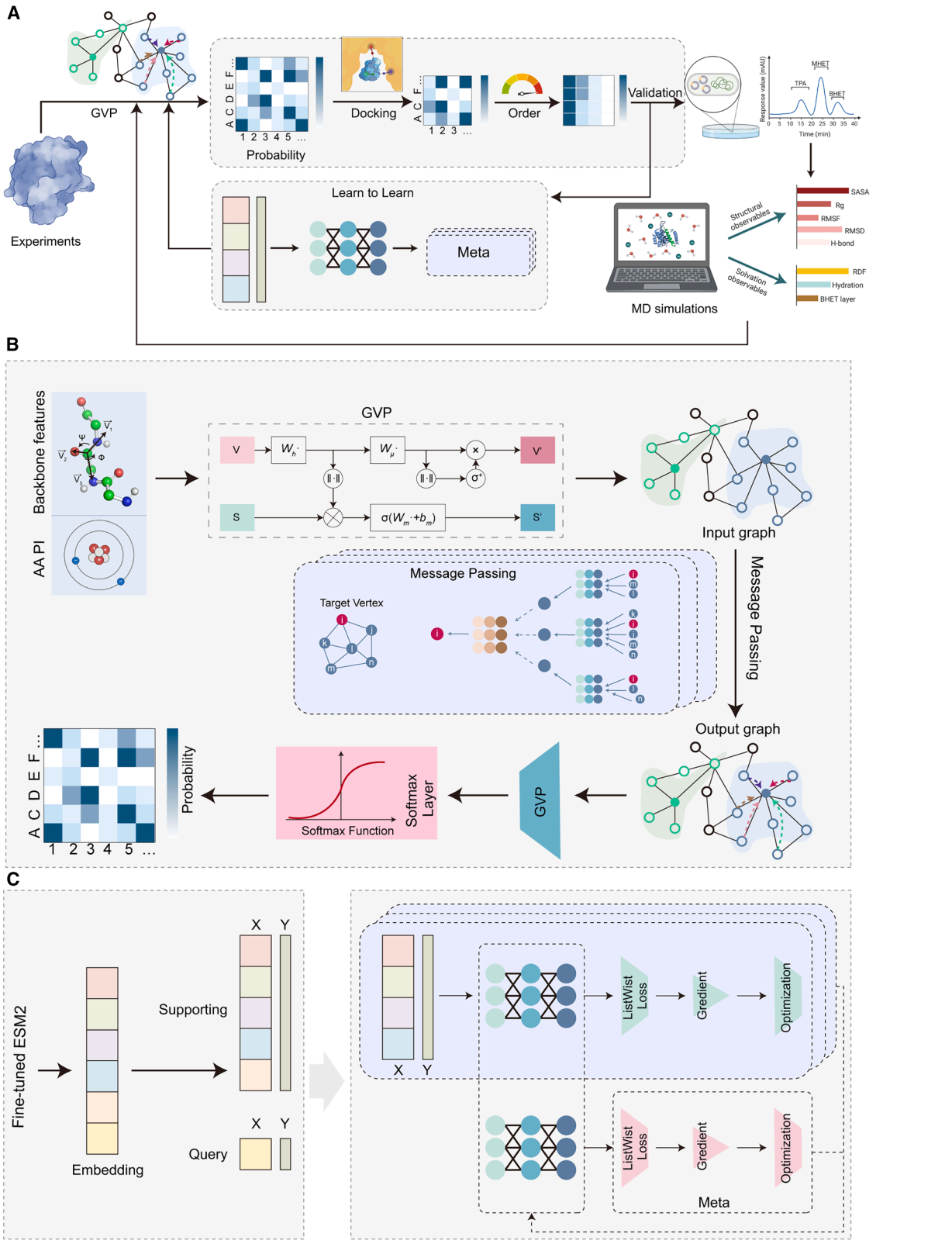
Among the petroleum-based plastics, polyethylene terephthalate (PET) is one of the most widely produced synthetic polymers, with a global manufacturing capacity exceeding 30 million tons annually. The extensive use of PET in packaging and textiles, combined with its recalcitrant nature, has led to pressing challenges in plastic waste management and environmental pollution. Although considerable progress has been made in developing biodegradable plastics,[1–4] their market presence remains limited compared to conventional nondegradable plastics.[2,5,6] Meanwhile, enzymatic PET degradation has also been intensively studied using engineered hydrolases, some of which have demonstrated activity at preparative or industrially relevant scales.[7–12] Given these limitations, biological recycling through enzymatic PET hydrolysis has garnered significant interest, and various PET hydrolases, including *Is*PETase,[13] Tfh,[14] leaf-branch compost cutinase (LCC),[15] Thc_Cut1,[16] Thh_Est,[17] and Hic,[18] have been biochemically and structurally characterized as potential candidates for PET depolymerization.

Efforts to enhance the efficiency and thermostability of PET hydrolases have led to notable breakthroughs by protein engineering campaigns.[7,8,19–21] In 2020, an engineered variant of LCC (LCC-ICCG) demonstrated 90% depolymerization of pretreated PET waste at high PET concentrations (200 g kg^{-1}),[12] and recent works have extended PETase and LCC variants to gram-scale or continuous processes.[7–12] Further advancements were achieved with DepoPETase, a highly engineered *Is*PETase variant (T88I/D186H/D220N/N233K/N246D/R260Y/S290P) produced via directed evolution, capable of depolymerizing untreated PET wastes in a liter-scale reactor (0.4% W_{enzyme}/W_{PET}).[20] Both LCC-ICCG and DepoPETase present robust scaffold candidates for further engineering aimed at enhancing thermostability and catalytic activity across variable temperature ranges.

Recent advancements in computational redesign have paved the way for significant progress in enzyme engineering.[22] Leveraging machine-learning (ML) and mutagenesis technologies has enabled the rapid and cost-effective exploration of protein structures and functions. Through large-scale biological data processing, deep-learning algorithms can decipher complex protein sequence and structural data, effectively predicting beneficial variants.[23] For instance, the evolutionary scale modeling (ESM)-1v protein language model developed by

CellPress
OPEN ACCESS

Cell Reports
Physical Science
Article



(legend on next page)

Please cite this article in press as: Li et al., Integrating graph learning and meta-learning to enhance PET hydrolase activity at elevated temperatures, Cell Reports Physical Science (2025), https://doi.org/10.1016/j.xcrp.2025.103037

Facebook AI, based on Transformer architecture, allows for unsupervised learning from vast protein sequence data, predicting variants that outperform the wild type in stability or activity.[24] ML-guided enzyme engineering circumvents some limitations of directed evolution, such as high screening costs and extended experimental cycles, by directly providing high-confidence predictions on enzyme modifications. Recent studies, including those by Alper's group[7] using MutCompute and by Hu et al.[25] using Bayesian optimization, have demonstrated the efficacy of ML in identifying variants with enhanced activity and stability through targeted predictions. Alper's team used MutCompute, a structure-based convolutional neural network (CNN), to a stabilizing five-site variant (FAST-PETase, N233K/R224Q/S121E/D186H/R280A) that exhibited superior PET-hydrolytic activity between 30°C and 50°C and a range of pH levels.[7] Besides this, Hu et al.[25] used Bayesian optimization to guide directed evolution, combining the sampling process with iterative optimization of the proxy model, and successfully obtained a four-site RhlA variant that could increase the product selectivity by 4.8-fold after four rounds of iterations. These developments underscore the emerging trend of digital protein design in synthetic biology.[26]

However, challenges remain in designing PET hydrolases through ML, primarily due to the complexity of balancing enzyme thermostability, catalytic efficiency, and substrate affinity.[7,27,28] Current ML strategies often focus on limited functional metrics (such as activity or stability), overlooking the potential trade-offs in evolving enzymes with optimal functionality across diverse and often conflicting properties.[29,30] Additionally, existing models are typically optimized based on available datasets with limited diversity, leading to predictions that may not generalize well to novel enzyme scaffolds. Consequently, the predictive power for functional and stable PET hydrolase variants that operate effectively at high temperatures remains constrained. To address these issues, this study presents a multi-tiered ML strategy that combines a geometric vector perceptron (GVP)-graph neural network (GNN) model with iterative meta-learning refinement through Gaussian and learn2learn models. The two-round approach allows for a broader exploration of mutational landscapes in the initial stage, followed by targeted refinement based on the identified "best" variants (Figure 1A). This strategy addresses several challenges in enzyme design, including (1) achieving a more comprehensive search of functional variants, (2) balancing thermostability with catalytic performance, and (3) accommodating diverse structural scaffolds of PET hydrolases. Additionally, we introduce an *in silico* scoring criterion to evaluate enzyme robustness comprehensively and systematically across different substitutions. Using molecular dynamics (MD) simulations (Figure 1A), we also investigated PET degradation to elucidate the mechanisms underlying hydrolytic activity at elevated temperatures and the microbehavior of intermediates (bis-2-hydroxyethyl terephthalate [BHET] and mono-2-hydroxyethyl terephthalate [MHET]). This alignment of *in vitro* and *in silico* studies not only underscores the utility of computational strategies in enzyme engineering but also highlights the importance of a two-enzyme system for efficiently producing the terminal product, terephthalic acid (TPA), thereby contributing to a sustainable solution for PET waste recycling.

RESULTS AND DISCUSSION

Development of a deep-learning framework for engineering PET hydrolases

In this study, we applied deep-learning techniques, specifically graph learning and meta-learning, aiming to narrow the candidate variant space and focus on residues at the enzyme-substrate interaction interface. The effectiveness of this comprehensive framework was validated on three representative protein scaffolds of PET hydrolases (two PETases: FAST-PETase and DepoPETase; and a BHETase: BsEst), the latter serving as a critical catalyst in PET biodepolymerization. In the first phase, we employed the optimized (GVP-GNN) model. This model enhances the ability to learn geometric relationships between nodes by incorporating geometric vector representations into the GNN framework. Each amino acid is treated as a node in the graph, with edges representing interactions between residues. To better capture spatial structure, dihedral angles, electrostatic charges, and other microenvironmental properties, we augmented node features with amino acid isoelectric point (pI) information[31] (Figure 1B). We assigned the pI value of each amino acid residue using its theoretical value derived from the Bjellqvist pK_a scale,[32] as retrieved from publicly available biochemical databases. These standard values were uniformly applied across all residues to maintain computational efficiency. We acknowledge that fixed pI values cannot fully represent charge rearrangements occurring during substrate interactions. This approximation was selected to ensure computational efficiency, while more rigorous treatments would require significantly greater computational resources. Each pI value was encoded as a scalar node feature and integrated with other

Figure 1. Graph-learning- and meta-learning-guided two-round predictions improve enzyme performance

(A) Workflow of enzyme performance improvement. GVP-GNN was adapted to predict mutants by assessing the difference between the original type of amino acid and the predicted type with the highest probability at a given site. Mutants closer to the active site or ligand (after docking) were more likely to be selected as experimental candidates. Based on the results of the first-round experiments, meta-learning was utilized to predict second-round mutants. Subsequently, molecular dynamics (MD) simulation was employed to understand the microscopic mechanism.

(B) Structural features and pI are input into a graph following transformations by GVP layers. The input graph undergoes further transformation through message-passing layers, influencing the embedding of residues based on their neighboring node and edge embedding. Finally, the node embeddings from the output graph are employed for mutant prediction, relying on probability and undergoing another GVP transformation.

(C) The embedding, derived from a fine-tuned ESM2 model and first-round experimental data, is divided into support and query sets denoted as X and Y, respectively. The following steps are repeated for all episodes of the meta-training dataset until the learner converges to a good set of meta-parameters: (1) sample a batch of episodes from the meta-training dataset; (2) train the adapter based on the support sets of the batch, compute the listwise loss and gradients, and update the adapter's parameters; (3) use the updated parameters of the adapter to compute the meta-loss based on the query sets of the batch; and (4) compute the meta-gradients, followed by the meta-parameters based on the meta-loss, and update the learner's parameters.

Please cite this article in press as: Li et al., Integrating graph learning and meta-learning to enhance PET hydrolase activity at elevated temperatures, Cell Reports Physical Science (2025), https://doi.org/10.1016/j.xcrp.2025.103037

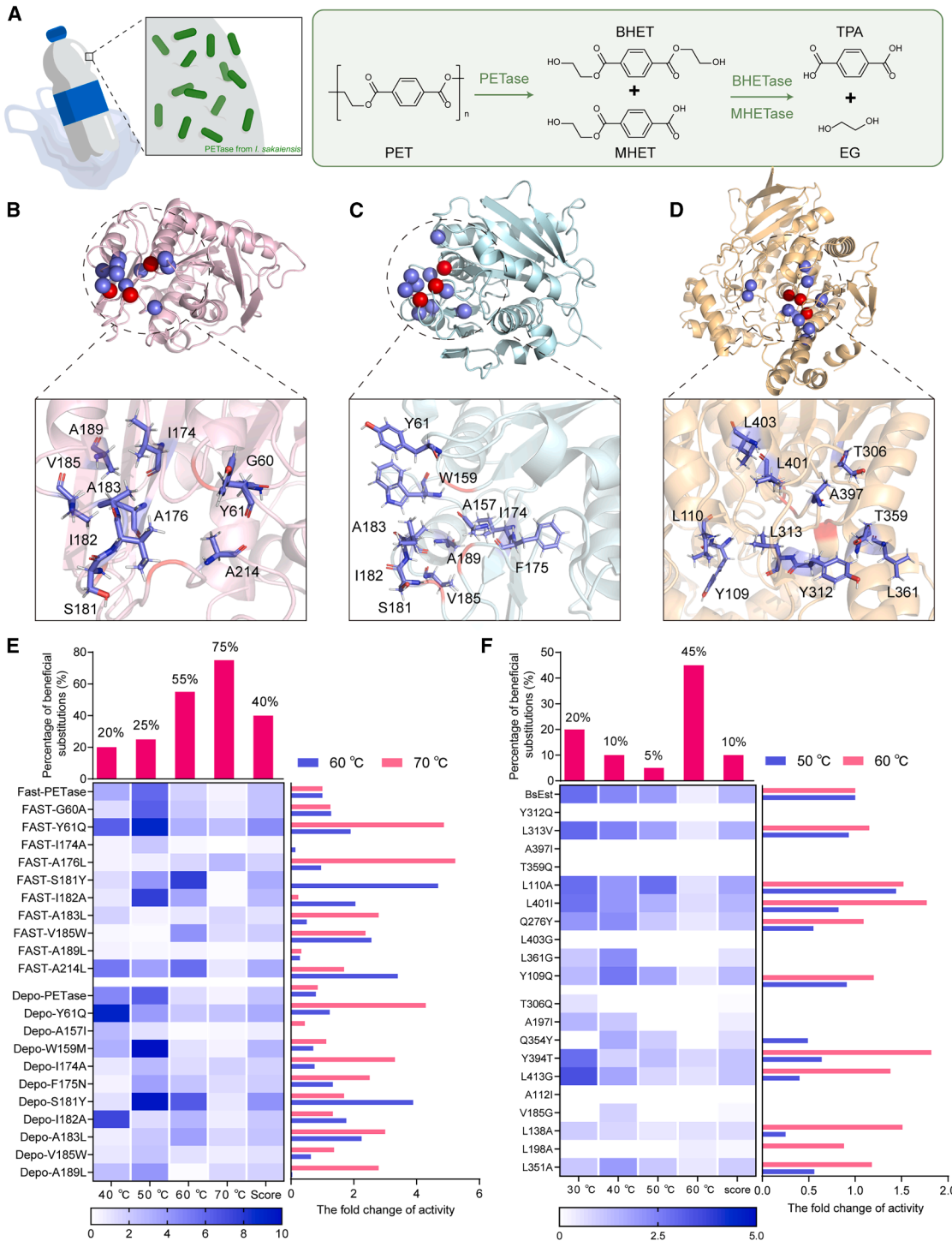


Figure 2. Machine-learning-guided predictions improve enzyme performance across AST-PETase, DepoPETase, and BsEst scaffolds

(A) Biodegradation mechanisms of PET. PET, polyethylene terephthalate; TPA, terephthalic acid; EG, ethylene glycol; BHET, bis-2-hydroxyethyl terephthalate; MHET, mono-2-hydroxyethyl terephthalate.

(B–D) Top 20 substitutions are distributed in the (B) FAST-PETase, (C) DepoPETase, and (D) BsEst protein structure by machine learning.

(E) The blue heatmap (left) shows the PET-hydrolytic activity of the resulting variants, and the histogram (right) shows the fold change of activity over FAST-PETase and DepoPETase scaffolds. The total PET monomers (the sum of TPA, MHET, and BHET) were released from hydrolyzing circular gf-PET film (6-mm-diameter pieces, roughly 67 mg) after 120 h of incubation at temperature ranging from 40°C to 70°C. Statistical normative scores calculated by assigning weights of 10%, 20%, 30%, and 40% to the TPA yield under 40°C, 50°C, 60°C, and 70°C, respectively.

(legend continued on next page)

Please cite this article in press as: Li et al., Integrating graph learning and meta-learning to enhance PET hydrolase activity at elevated temperatures, Cell Reports Physical Science (2025), https://doi.org/10.1016/j.xcrp.2025.103037

physicochemical properties as input to the GVP-GNN, enabling better representation of local electrostatic environments relevant to thermostability and catalysis. The microenvironment of the protein, including backbone and pI information, was concatenated and fed as input features to the GVP layer, forming the initial input graph. Using a message-passing mechanism, nodes iteratively updated their feature representations by interacting with their neighbors, allowing the model to capture complex interdependencies among residues. The final output predicted the probability distribution of 20 possible amino acids for each site. Additionally, we integrated predicted variant probabilities with spatial information, such as the distance between candidate residues and the specific substrate molecule. This combined analysis allowed us to select the top ten candidates for experimental validation. We note that ML models may exhibit prediction bias (over- or underestimation) in absolute activity values. In this work, the models were primarily used for relative ranking of variants, with experimental assays providing validation and feedback for iterative refinement in the design-test-learn cycle.

In the second phase, we employed a meta-learning strategy, which excels in handling small-sample data compared to traditional methods like learn2learn and Gaussian processes. For BsEst, experimental data at multiple temperatures were treated as separate tasks, each containing a support set and a query set. A pretrained protein language model based ESM2 was utilized for meta-learning to better fit enzyme activity data with limited samples. In detail, to enhance the sequence embedding of BsEst, we used hhblits to retrieve homologous sequences and fine-tuned the protein language model ESM2 using this sequence set. The fine-tuned ESM2 embeddings were then divided into support and query sets for model diagnosis through meta-learning using the model-agnostic meta-learning (MAML) framework (Figure 1C). Our meta-learning strategy currently relies solely on sequence embedding (ESM2), which may limit cold-start performance. Future work integrating structural or phylogenetic features could further enhance predictive accuracy and broaden applicability.

GVP-GNN model to predict the round I variants for enhancing thermostability toward three PET hydrolases with two types of protein scaffolds

The highly focused protein engineering approaches usually cannot consider the evolutionary trade-off between overall stability and activity, and the last few years have witnessed impressive progress in tailoring natural enzymes by computational redesign strategies.[22] Specifically, the engineered FAST-PETase through CNN[7] and DepoPETase[20] by directed evolution showed marvelous capability in hydrolyzing PET into monomers (including BHET, MHET, TPA, and ethylene glycol [EG]) at high temperature (50°C–60°C, Figure 2A). Additionally, our previous study identified BHETase, which catalyzes the conversion of the intermediate product BHET into the terminal monomer TPA (Figure 2A). Based on these two promising PETase scaffolds (FAST-PETase and DepoPETase) and one BHETase scaffold (BsEst), we first employed the refined GVP-GNN model to further predict their extra tailorable positions (Figures 2B–2D). Discrete probability distributions for all 20 standard amino acids at each position across the three protein scaffolds were generated, effectively conducting an *in silico* integrated mutagenesis scan. The predictions were ranked based on their probabilities, and the top ten site distributions were illustrated on the 3D structure of the protein to identify locations where the wild-type amino acid residues poorly matched potential substitutions (Figures 2B–2D). We subsequently evaluated PET-hydrolytic activity of these variants across a range of temperatures from 40°C to 70°C using an amorphous PET film (gf-PET, from the supplier Goodfellow).[33] An *in silico* scoring criterion was introduced to comprehensively assess the robustness of these variants (Figures 2E and 2F). For each variant, we quantified TPA production at four representative temperatures (40°C, 50°C, 60°C, and 70°C for FAST-PETase and DepoPETase). A normalized score was then calculated as a weighted sum of TPA concentrations using the formula

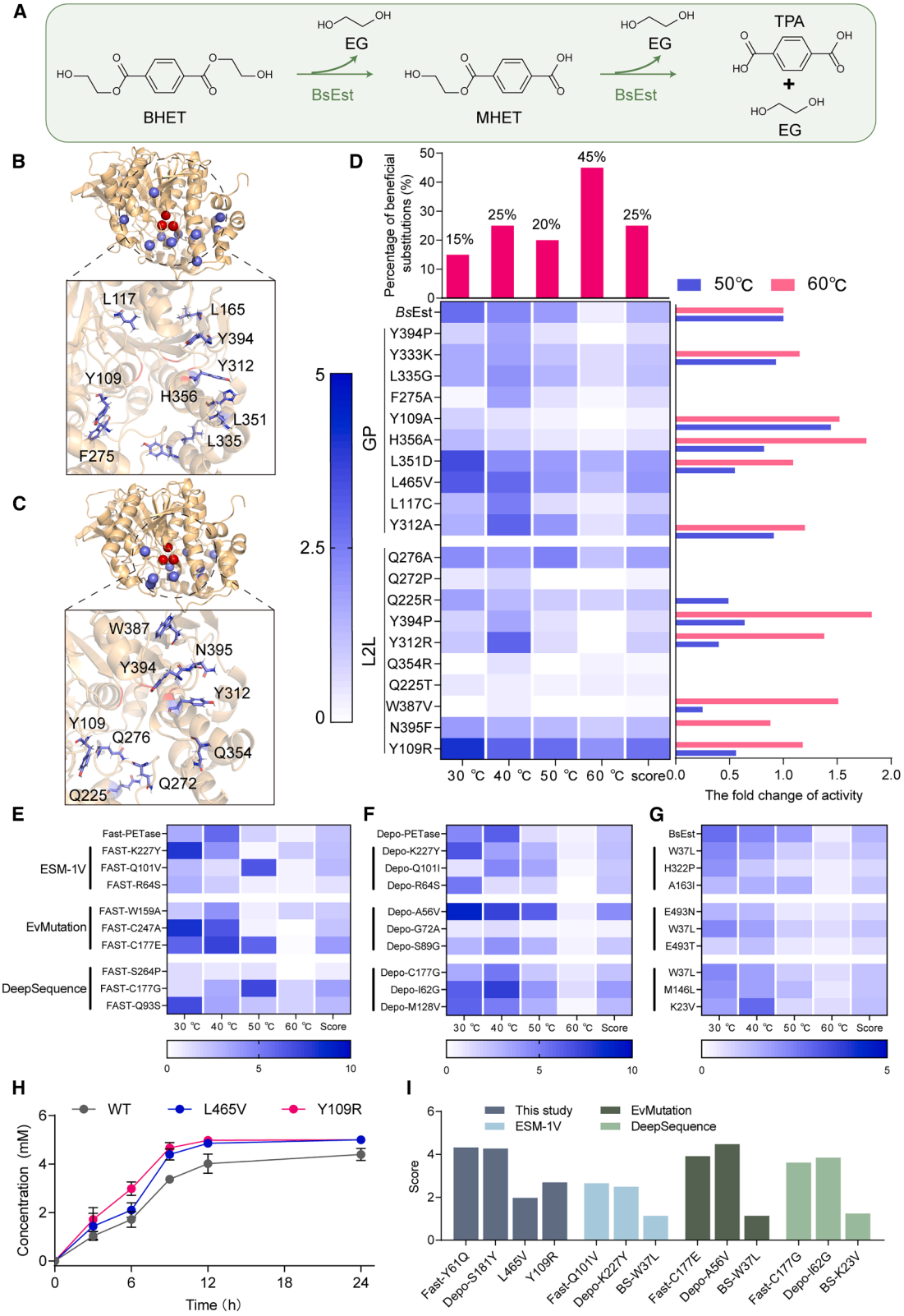
$$\text{Normalized score} = (C_{40} \times 10\%) + (C_{50} \times 20\%) + (C_{60} \times 30\%) + (C_{70} \times 40\%),$$

where C_x denotes the TPA concentration (mM) at temperature x. This weighting reflects the greater relevance of high-temperature activity in industrial PET degradation. A similar strategy was used for BHETase variants, evaluated across 30°C–60°C.

Overall, the beneficial rate is ranging from 20% to 75% under a range of investigated temperatures from 40°C to 70°C (Figure 2E, top). Among the predicted top ten variants, FAST-Y61Q demonstrated a 1.9-fold and 4.9-fold increase in PET-hydrolytic products at 60°C and 70°C, respectively. Aligning with the *in silico* scoring criteria, FAST-Y61Q (score: 4.32) significantly outperformed the FAST-PETase scaffold (score: 2.17). Variants based on the DepoPETase scaffold exhibited an extended operational temperature range of 40°C–70°C, which was generally higher than that of the FAST-PETase scaffold variants (Figures 2E and S1A). The best-performing variant from the DepoPETase scaffold was DepoPETase-S181Y, which achieved normalized scores of 4.27, resulting in 1.6-fold and 4.9-fold higher activity at 50°C and 60°C, respectively. The crude enzymes of FAST-Y61Q and DepoPETase-S181Y showed the highest overall degradation of all variants, and temperatures tested at 50°C and 60°C released 7.63 mM and 9.56 mM PET monomers (the sum of BHET, MHET, and TPA), respectively, over 96 h (Figure 2E). In addition, the reaction progression profiles of both FAST-Y61Q and Depo-S181Y were monitored at 60°C (Figures S1C and S1D). To thoroughly evaluate the catalytic adaptation of the variant enzymes under varying environmental conditions, FAST-Y61Q, Depo-S181Y, and their respective protein scaffolds were assessed across a pH range of 5.5–8.5 using amorphous gf-PET (Figure S2). These results underscore that GVP-GNN-guided predictions significantly enhanced enzyme

(F) The blue heatmap (left) shows the terminal TPA production of the resulting variants, and the histogram (right) shows the fold change of activity over the BsEst scaffold. TPA production released from hydrolyzing 5 mM BHET by the BsEst variants after 6 h of incubation at temperature ranging from 30°C to 60°C. Statistical normative scores calculated by assigning weights of 10%, 20%, 30%, and 40% to the TPA yield under 30°C, 40°C, 50°C, and 60°C, respectively.
All measurements were conducted in triplicate (n = 3), and the mean values were used for generating heatmaps and histograms.

Please cite this article in press as: Li et al., Integrating graph learning and meta-learning to enhance PET hydrolase activity at elevated temperatures, Cell Reports Physical Science (2025), https://doi.org/10.1016/j.xcrp.2025.103037



(legend on next page)

activity while extending the operational temperature and pH range across all scaffolds. To benchmark the performance of our designed enzymes, we compared the representative variants FAST-Y61Q and DepoPETase-S181Y with the state-of-the-art TurboPETase and HotPETase under identical conditions. As shown in Figure S4, both variants exhibited comparably higher TPA release after 48 h, demonstrating that our ML-guided design can generate competitive or superior PET-degrading enzymes relative to existing high-performance PETases.

Next, we sought to investigate the generalizability of our GVP-GNN method by investigating another protein scaffold, BHETase. Although the beneficial rate ranges from 5% to 45% regarding the investigated temperatures (Figure 2F, top), the top ten BsEst variants identified by the GVP-GNN model showed no significant enhancement in catalytic activity at temperatures ranging from 30°C to 60°C. The L110A variant, which had the highest normalized score (1.74), only achieved a 1.2-fold increase compared to wild-type BsEst (1.45). While L110A produced 1.5-fold increased TPA at 60°C and demonstrated the highest TPA production at 30°C and 50°C, this performance was still below our expectations. Subsequently, we expanded the recommended range of the GVP-GNN model to include the top 20 variants, finding that the next ten variants had significantly lower comprehensive scores than the top ten (Figure 2F). Given these results, we prioritized further optimization of BsEst through the meta-learning strategy (Gaussian model and learn2learn) to enhance its performance. This decision allowed us to efficiently allocate experimental resources to the scaffold that most required improvement while also demonstrating the adaptability and complementarity of the two-round ML strategy.

Meta-learning was used for round II predictions to identify extra robust BHETase variants

Based on the experimental results from round I, where the GVP-GNN model recommended the top 20 variants, we employed meta-learning with both the Gaussian model and the learn2learn model to fit the respective experimental results in Figure 2. We then evaluated the hydrolytic activity of BHET at temperatures ranging from 30°C to 60°C for the top ten variants recommended from each model, as illustrated in Figures 3B and 3C. Notably, Figure 3D demonstrates that the number of beneficial variants identified in round II using the Gaussian model and the learn2-learn model significantly exceeded those identified in round I with the GVP model. Among these variants, the BsEst L465V, which achieved a normalized score of 1.97 from the Gaussian model, exhibited a 3.1-fold increase in terminal product TPA compared to the wild type at 60°C. Meanwhile, the BsEst Y109R variant, which received the highest score of 2.69 from the learn2learn model, surpassed the wild-type BsEst score of 1.45. Interestingly, Y109R demonstrated superior activity across a range of temperatures (30°C–60°C), with a TPA yield that was 5.5-fold greater than that of the wild type at 60°C.

Encouraged by these promising results, we monitored the BHET hydrolysis progress at a high temperature of 60°C using the purified Y109R and L465V variants. As depicted in Figures 3H and S3, both variants achieved complete conversion of BHET within 9 h and 100% TPA yield within 12 h. Notably, the rate of TPA production by Y109R (0.5 mM/h) was higher than that of L465V (0.35 mM/h).

Next, we compared three state-of-the-art ML algorithms—ESM-1V,[24] EvMutation,[34] and DeepSequence[35]—with the algorithms used in this study (Figures 3E–3G). Since we explored only 10–40 variants for each enzyme, and BsEst Y109R and FAST-Y61Q consistently ranked among the top three predicted variants, we focused our efforts on experimentally validating the top three variants identified by three state-of-the-art ML algorithms. It should be noted that comparisons with other models were limited to the top three predictions due to experimental cost constraints and thus were intended to provide context rather than to establish rigorous performance ranking. Collectively, the number of beneficial variants at medium temperatures (30°C–50°C) was notably higher than at high temperatures (60°C–70°C). To systematically identify beneficial variants, we assigned varied weights (10%, 20%, 30%, and 40%) to gradient temperatures and tallied the optimal variant scores from each algorithmic model (Figure 3I). In round II, the meta-learning strategy leveraged episodic sampling and a ranking-based loss (ListMLE) to prioritize stable BHETase variants from limited training data, distinguishing it from conventional regression-based approaches. Our method systematically identified superior beneficial variant scores across all models, outperforming ESM-1V, EvMutation, and DeepSequence, particularly for BHETase (Figure 3I). While our framework generally outperformed baseline models, we note that EvMutation achieved comparable or superior predictions in certain cases. This likely reflects the strength of evolutionary constraints embedded in EvMutation, which are especially informative for mutations occurring at highly conserved positions.[34] These results collectively demonstrate that the ML approach proposed in this study, combined with *in silico* scoring criteria, is most likely to address the persistent challenges associated with engineering robust PET hydrolases. Owing to the dense molecular packing and limited

Figure 3. Machine-learning-guided predictions improve enzyme performance across BsEst scaffolds

(A–C) (A) Biodegradation mechanisms of BHET. Top ten substitutions are distributed in the wild-type BsEst protein structure by (B) Gaussian model and (C) learn2learn model.

(D) The blue heatmap (left) shows the terminal TPA production of the resulting variants, and the histogram (right) shows the fold change of activity over the BsEst scaffold. TPA production released from hydrolyzing 5 mM BHET by the BsEst variants after 6 h of incubation at temperature ranging from 30°C to 60°C. Statistical normative scores calculated by assigning weights of 10%, 20%, 30%, and 40% to the TPA yield under 30°C, 40°C, 50°C, and 60°C, respectively. All measurements were conducted in triplicate (n = 3), and the mean values were used for generating heatmaps and histograms.

(E–G) Algorithmic comparison of ESM-1V model, EvMutation model, and DeepSequence model based on the (E) FAST-PETase, (F) DepoPETase, and (G) BsEst scaffolds.

(H) Biodegradation performance of TPA production from BsEst upon incubation on BHET. Reaction conditions: 5 mM BHET at 30°C in buffer (pH 7.5) for 24 h. Error bars correspond to the standard deviation (SD) of three measurements (n = 3).

(I) The highest-scoring variants were counted under four algorithms.

Please cite this article in press as: Li et al., Integrating graph learning and meta-learning to enhance PET hydrolase activity at elevated temperatures, Cell Reports Physical Science (2025), https://doi.org/10.1016/j.xcrp.2025.103037

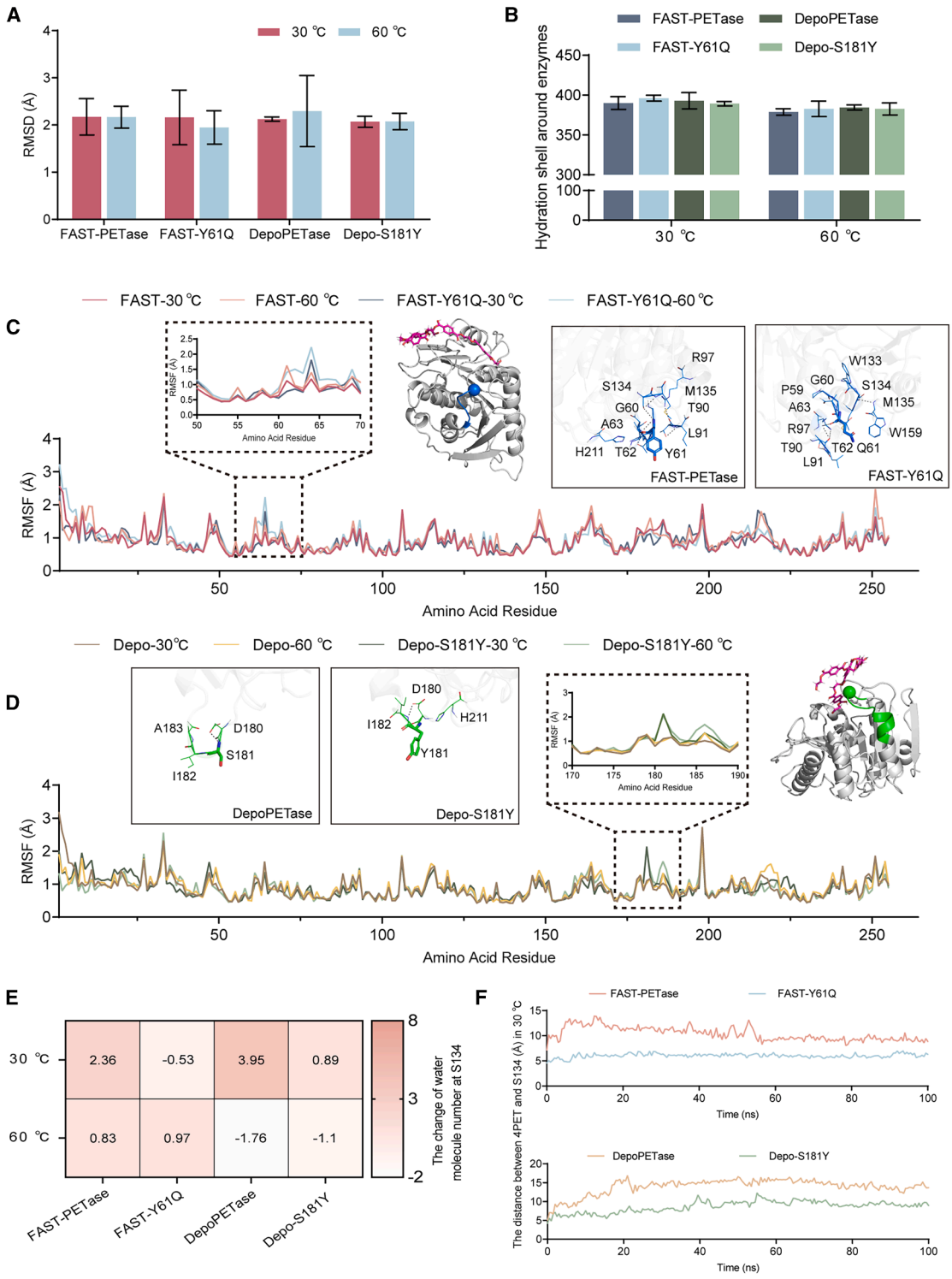


Figure 4. Molecular insight into the overall structure of FAST-PETase and DepoPETase variants

(A) The time-averaged RMSD of the backbone atoms (C_α, N, and carbonyl C) of the FAST-PETase, DepoPETase, and their variants with respect to the initial structure at 30°C and 60°C, respectively, determined from the last 40 ns of MD simulation.

(B) The number of hydration shells of the FAST-PETase, DepoPETase, and their variants at 30°C and 60°C, respectively, during last 40 ns simulation from three independent runs. The hydration shell was defined as water molecules whose oxygen atom was localized at a distance less than 3.5 Å from any nonhydrogen atom of the enzyme.

(legend continued on next page)

Please cite this article in press as: Li et al., Integrating graph learning and meta-learning to enhance PET hydrolase activity at elevated temperatures, Cell Reports Physical Science (2025), https://doi.org/10.1016/j.xcrp.2025.103037

chain mobility of highly crystalline PET, most PET hydrolases are unable to efficiently catalyze depolymerization directly.[36,37] Complete degradation of high-crystallinity PET substrates, therefore, typically requires additional chemical or physicochemical pretreatment steps—such as thermal annealing, mechanical milling, or solvent swelling—to increase polymer accessibility and create more amorphous regions.[38–40] Integrating such pretreatment strategies with the use of high-performance, ML-screened enzyme variants offer a promising route toward efficient and scalable PET recycling. The synergistic combination of substrate activation and catalyst optimization could substantially lower energetic barriers, enhance hydrolysis kinetics, and enable the conversion of otherwise recalcitrant PET into its monomeric building blocks under mild conditions. This integrative approach not only bridges material and enzyme engineering but also lays the groundwork for a closed-loop sustainable plastic upcycling process.

Unraveling the structural basis for enhanced thermostability

To investigate the effect of high temperature on enzyme activity, we conducted MD simulations for wild-type FAST-PETase, DepoPETase, BsEst, and their identified variants. This approach allowed us to examine structural and solvation phenomena from a molecular perspective (Figure 4). During the simulations, we closely monitored the root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) of the backbone atoms, including C_α, N, and carbonyl C, to assess thermal fluctuations in protein conformation (Figures 4A, 4C, 4D, S5, and S8). The RMSD values indicated that the four variants (FAST-Y61Q, Depo-S181Y, BsEst-L465V, and BsEst-Y109R) exhibited minimal changes of less than 0.3 Å, even at 60°C (Figures 4A and S5). Notably, FAST-Y61Q and Depo-S181Y displayed higher RMSF fluctuations observed at the Y61 site of FAST-PETase and the S181 site of DepoPETase (~1 Å) but with similar RMSF trends within substrate-binding cleft regions (Figures 4C and 4D). This observation correlated with a decrease in the number of hydrogen bonds around these variant sites (Figures 4C and 4D). In proteins, hydrogen bonds are typically defined by a hydrogen–acceptor distance of less than 2.5 Å and a donor-hydrogen-acceptor angle between 90° and 180°, which reflects the most commonly accepted geometric criteria.[41] These results suggest that substitutions at Y61 and S181 induce significant structural fluctuations in localized regions but have a limited impact on the overall enzyme structure, indicating that the catalytic function of the variants enables them to be maintained even at elevated temperatures. The maintenance of a sufficient number of hydrogen bonds was evidenced by calculations of numbers of internal hydrogen bonds in overall and local regions along with a complex network of hydrogen-bonding interactions around the substitution sites (Figures 4C, 4D, and S9–S11). The latter plays a crucial role in preserving structural stability. Further details regarding the structural analysis from the MD simulations can be found in Figures S5–S11 and Note S1. In addition to hydrogen-bonded stabilization, we analyzed the *B* factors of the top variants (Figure S5) to evaluate residue flexibility. Several sites, including FAST-Y61, FAST-S181, FAST-I182, FAST-A214, and Depo-Y61, displayed relatively higher *B*-factor values, implying increased local dynamics that may facilitate substrate accommodation. These results indicate that our design strategy balances thermal stability with catalytic adaptability, aligning with the dual objective of enhancing both enzyme robustness and activity toward plastic substrates.

Additionally, the solvation phenomena observed through MD simulations clearly demonstrated the driving force behind the enhanced degradation capacity observed between the model substrates 4PET/BHET and their variants.[43,44] The hydration shell—defined as water molecules with oxygen atoms located within 3.5 Å of any non-hydrogen atom of the enzyme[41]—remained largely unchanged in the four variants (FAST-Y61Q, Depo-S181Y, BsEst-L465V, and BsEst-Y109R) compared to their respective wild-type scaffolds (Figures 4B and S12). The number of water molecules around the active sites of FAST-Y61Q and Depo-S181Y (3.58 and 5.02) was significantly lower than that of FAST-PETase and DepoPETase (6.47 and 8.06) at 30°C. Less space occupied by water molecules in the active pocket facilitates the entry of the hydrophobic substrate 4PET/BHET into the vicinity of the catalytic triad and promotes the hydrolysis reaction (Figures 4E and S14). These results are aligned with the observed changes in the distance between the nucleophilic Ser and substrate 4PET over the 100 ns simulation period within the catalytic triad (Figures 4F, S15, and S16). However, at high temperature (60°C), the number of water molecules did not show a significant difference, and the reaction distance exhibited a corresponding result (Figure S14), suggesting that there may be other reasons promoting the catalytic efficiency of the variants, such as the conformational change of the active pockets, and the main reason may be the enhanced thermostability of the overall structure of the variants analyzed above (Figure 4A), which maintains its own activity. The increased

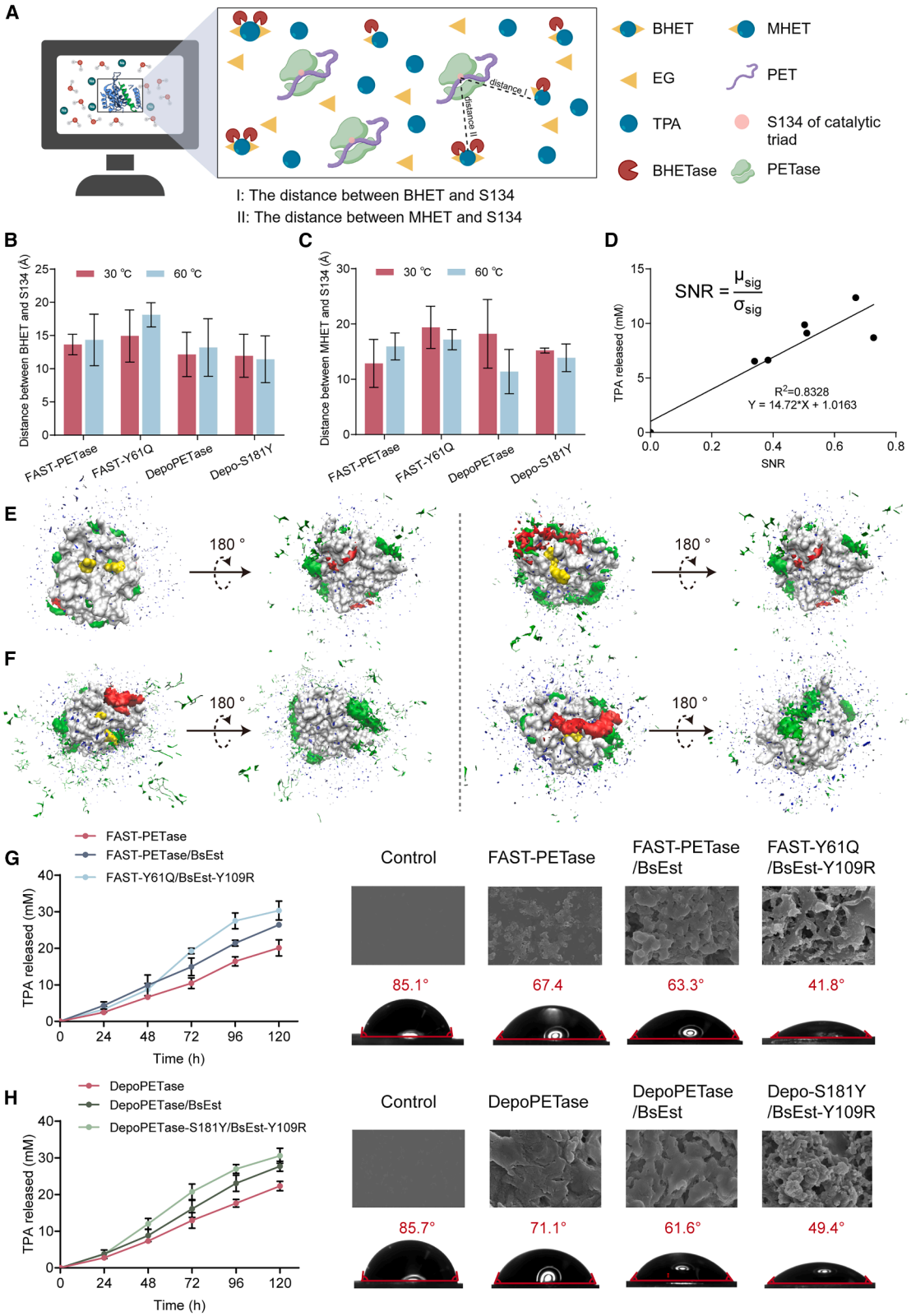
(C) RMSF of the wild-type FAST-PETase and FAST-Y61Q variant at 30°C and 60°C, respectively, determined from the last 40 ns of MD. The pink stick represents the model substrate 4PET; the blue sphere represents the substitution of Y61; and the blue loop represents the regions of large fluctuations in RMSF values. Hydrogen-bond network of FAST-PETase and FAST-Y61Q within 4.6 Å around Y61. Residues within 4.6 Å are presented as lines, Y61 is presented as a stick, and both are colored blue.

(D) RMSF of the wild-type DepoPETase and Depo-S181Y variants at 30°C and 60°C, respectively, determined from the last 40 ns of MD. The pink stick represents the model substrate 4PET; the green sphere represents the substitution of S181; the green loop represents the regions of large fluctuations in RMSF values. Hydrogen-bond network of DepoPETase and Depo-S181Y within 3.55 Å around S181. Residues within 3.55 Å are presented as lines, S181 is presented as a stick, and both are colored green. Data plotted from the average of three independent MD runs (n = 3).

(E) Number of water molecules around the active site (S134) of the FAST-PETase, DepoPETase, and their variants at 30°C and 60°C, respectively. The distance around S134 was delineated as 6.6 Å, which was predicted by the DEPTH server (http://cospi.iiserpune.ac.in/depth).[42] Error bars correspond to the SD of three independent MD runs (n = 3).

(F) Distance curves of FAST-PETase, DepoPETase, and their variants between the hydroxyl oxygen atom (OG) of S134 and the carbonyl carbon atoms (C40) of 4PET at 100 ns simulation proceeding at 30°C.

Please cite this article in press as: Li et al., Integrating graph learning and meta-learning to enhance PET hydrolase activity at elevated temperatures, Cell Reports Physical Science (2025), https://doi.org/10.1016/j.xcrp.2025.103037



(legend on next page)

Please cite this article in press as: Li et al., Integrating graph learning and meta-learning to enhance PET hydrolase activity at elevated temperatures, Cell Reports Physical Science (2025), https://doi.org/10.1016/j.xcrp.2025.103037

number of water molecules at the mutation sites could enhance catalytic efficiency (Figure S14A). Collectively, the synergistic effects of structural stability and catalytic capacity resulting from the introduction of FAST-Y61Q and Depo-S181Y contribute to an improved catalytic performance, especially at elevated temperatures (60°C). A similar conclusion was also reached for BsEst with the substrate BHET (Figures S5, S6B, S7C, S8, S9B, S10–S13, S14B, and S16). To enhance the reliability of the MD simulation data in this study, longer 1 μs MD simulations were conducted under identical conditions, yielding comparable results. Detailed analyses are provided in Note S1.

Construction of a two-enzyme system in PET depolymerization guided by the molecular behaviors of intermediates

In a recent study, Wang et al. employed high-level quantum mechanics/molecular mechanics (QM/MM) calculations, specifically density functional theory (DFT), to accurately determine reaction barriers and elucidate the catalytic mechanism of PETase.[45] PETase-catalyzed PET degradation involves four steps: (1) nucleophilic attack triggered by Ser-His-Asp, (2) cleavage of the C–O bond, (3) nucleophilic attack by water molecule, and (4) PETase deacylation. Notably, we observed that during the C–O bond cleavage step (step 2) and the subsequent PETase deacylation (step 4), intermediate products BHET and MHET are released from the active site.[45] This finding aligns well with the composition of PET degradation products detected using high-performance liquid chromatography (HPLC) as reported previously.[33] Experimental observations revealed that PET hydrolase (LCC) was predominantly collected from PET film rather than in solution, suggesting a strong affinity of PET hydrolase for binding to PET film.[46]

Based on insights derived from both computational calculations and experimental data, we proposed the construction of a dual-enzyme system to enhance the biodegradation of PET (Figure 5A). Prior to implementing this system, spatial distribution function (SDF) calculations were performed to analyze the arrangement of water molecules in the BHET and MHET intermediates. Simulations over 100 ns indicated that BHET and MHET remained solubilized, as evidenced by a significant decrease in the number of surrounding water molecules (Figure S17). This observation was further corroborated by the distances between the intermediates (BHET and MHET) and the nucleophilic serine exceeding 10 Å (Figures 5B and 5C). Furthermore, it was noted that BHET and MHET faced challenges in accessing the catalytic triad due to their positioning on the protein surface (Figure 5E), and they were rapidly displaced from the enzyme surface when model 4PET occupied the active site (Figure 5F). Given these computational results, we decided to introduce BHETase (BsEst), an enzyme responsible for converting intermediates into the terminal product TPA, into the PETase single-enzyme system. The incorporation of BHETase resulted in a significant 1.32-fold increase in TPA yield compared to the single-enzyme system (FAST-PETase and DepoPETase, Figures 5G and 5H). Furthermore, the integration of both PETase (FAST-Y61Q and Depo-S181Y) and BHETase (Y109R), engineered through ML algorithms, yielded a dual-enzyme system that produced homogeneous TPA with a 1.5-fold increase in yield. Notably, the rate of TPA generation in this enhanced two-enzyme system was up to 1.9 times greater than that of the previous system, measured over 48 h.

These findings were further substantiated by scanning electron microscopy (SEM) analysis and water contact angle measurements, which together provided supportive evidence of surface erosion and increased hydrophilicity of the PET films (Figures 5G and 5H). While more detailed bulk analyses such as size-exclusion chromatography or differential scanning calorimetry could further characterize polymer degradation, our current study focuses primarily on evaluating the catalytic performance of the engineered enzymes. The observed morphological changes and surface wettability shift correlate well with the enhanced TPA release, offering reliable indirect indicators of enzyme-mediated depolymerization. Specifically, the two-enzyme system demonstrated a remarkable capacity to effectively degrade the PET film, resulting in a reduction of the water contact angle by 25.6° and 21.7°, respectively, and producing a visibly rougher surface. Additionally, calculations of the signal-to-noise ratio (SNR) through image analysis revealed a correlation with the amount of released TPA from both the single- and two-enzyme systems ($R^2 > 0.8$, Figure 5D). This collective alignment of molecule-behavioral observations with experimental data robustly supports the finding that the dual-enzyme system plays a pivotal role in enhancing PET biodegradation. Nonetheless, further exploration of the quantitative relationships between SEM characterization and degradation effects is warranted to develop a novel evaluation system for PET degradation.

Figure 5. Two-enzyme degradation system including PETase and BHETase

(A) An overview of the two-enzyme degradation system. Icon graphics of this figure were created by BioRender.com.

(B) Average distances of FAST-PETase, DepoPETase, and their variants between the hydroxyl oxygen atom (OG) of S134 and the carbonyl carbon atom (C10) of ten BHET molecules at 30°C and 60°C, respectively, during 100 ns of MD.

(C) Average distances of FAST-PETase, DepoPETase, and their variants between the hydroxyl oxygen atom (OG) of S134 and the carbonyl carbon atom (C8) of ten MHET molecules at 30°C and 60°C, respectively, during 100 ns of MD.

(D) Linear regression between signal-to-noise ratio (SNR) and TPA released. The SNR is defined as the ratio of the average image signal value μ_{sig} to the SD of the image signal σ_{sig}.

(E and F) Spatial distribution function (SDF) of intermediate BHET and MHET on the surface of FAST-Y61Q and Depo-S181Y at 60°C when (E) 4PET was free in the system and (F) 4PET occupied the active site in a docking pose. Red represents the model substrate 4PET; green represents the intermediate productions including BHET and MHET; blue represents water molecules; and yellow represents the active site (S134). The 180° rotation offered to give a complete view of FAST-Y61Q and Depo-S181Y.

(G and H) Time course of PET depolymerization in a two-enzyme system with different PET hydrolases, including (G) PETase and (H) DepoPETase, and the corresponding SEM images (upper panels) and water contact angle analysis (lower panels) of the PET film in the two-enzyme degradation system. Error bars correspond to the SD of three independent MD runs ($n = 3$).

Please cite this article in press as: Li et al., Integrating graph learning and meta-learning to enhance PET hydrolase activity at elevated temperatures, Cell Reports Physical Science (2025), https://doi.org/10.1016/j.xcrp.2025.103037

This study presents an integrated ML framework, incorporating a two-round approach (round I: GVP-GNN model; round II: Gaussian model and learn2learn model) to comprehensively identify beneficial variants of PETase and BHETase by only exploring such a small library (10–40 variants). By applying *in silico* scoring criteria, we successfully engineered robust variants from the FAST-PETase and DepoPETase scaffolds, namely FAST-Y61Q and Depo-S181Y, as well as from the BHETase scaffold, including BsEst-L465V and BsEst-Y109R, demonstrating enhanced catalytic activity and functional stability across a broad temperature range (40°C–70°C), as supported by experimental assays and MD simulations. Importantly, the ML framework in this study outperformed existing models such as ESM-1V, EvMutation, and DeepSequence in identifying beneficial variants with the highest score, particularly for BHETase. Despite its success, our ML framework has certain limitations. For instance, it may not perform optimally for highly complex proteins that lack high-confidence structural data or detailed active-site information. Additionally, the framework's second stage (meta-learning with Gaussian and learn2learn) currently relies solely on sequence-based protein characterization models. Incorporating structural information in this stage could be a promising direction for future enhancements, potentially improving the framework's predictive power and applicability to a broader range of proteins. Besides, MD simulations offered molecular insights into the behaviors of intermediates, confirming that the construction of a dual-enzyme system effectively accelerates the homogeneous production of TPA. Collectively, this innovative synergy between the ML framework and the dual-enzyme system significantly converts PET hydrolase (PETase and BHETase) scaffolds into broad-range biocatalysts. These findings offer a proof of concept for scalable PET hydrolase engineering and highlight the potential of ML-assisted strategies to improve monomer yields under industrially relevant conditions. This study demonstrates the strength of meta-learning in enzyme engineering, enabling generalization to new conditions (e.g., 60°C) with limited data. Our framework achieved about 20% success in identifying beneficial variants from only 10–40 candidates per round, highlighting its efficiency and practicality. Although systematic ablation analyses were beyond the current experimental scope, future work will further refine and expand this paradigm.

METHODS

Chemicals, reagents, and materials

Substrates and reaction products, including BHET, TPA, and higher-purity chemicals, were purchased from Merck (Darmstadt, Germany) and Sigma-Aldrich (St. Louis, MO, USA), respectively. The amorphous PET film was purchased from Goodfellow (UK, product code: ES303010) and cut into 6-mm-diameter pieces (roughly 67 mg) for further experiments. Genes encoding FAST-PETase (*Is*PETase$^{D186H/R280A/N233K/R224Q/S121E}$, *Is*PETase GenBank: BBYR01000074) and DepoPETase (*Is*PETase$^{T88I/D186H/D220N/N233K/N246D/R260Y/S290P}$) were commercially synthesized by Tsingke Biotechnology (Nanjing, China) with codon optimization for expression in *Escherichia coli* BL21 cells. BsEst was identified in our previous study and stored in our lab.[33]

Site-directed mutagenesis

Variants of FAST-PETase, DepoPETase, and BsEst were constructed by PCR plasmid amplification. PCR products were incubated with DpnI (New England Biolabs, Ipswich, MA, USA) to digest the original DNA template, then separately transformed into *E. coli* DH5α (heat shock at 42°C for 45 s), followed by expressing in *E. coli* BL21(DE3). Finally, the introduced mutations were confirmed by sequencing (Tsingke Biotechnology).

Enzyme expression and purification

The recombinant bacteria were inoculated into Luria-Bertani (LB) liquid medium (100 μg/mL ampicillin) and grown overnight at 37°C/200 rpm for 14 h. The 2% cell cultures were transferred into 100 mL of LB liquid medium (100 μg/mL ampicillin) and shaken at 37°C/200 rpm for 3 h. When the optical density at 600 nm (OD_{600}) reached 0.6–0.8, isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to final concentrations of 0.1 mM and grown overnight at 20°C/200 rpm for 24 h. Subsequently, supernatants were collected by centrifugation (8,000 × *g* for 10 min at 4°C) and resuspended in 50 mM Tris-HCl buffer (pH 7.5) after washing twice with buffer (pH 7.5). The washed cells with 1 mg/mL lysozyme were disrupted by sonication (3-s pulse on, 5-s pulse off, amplitude 35%). Cell debris was removed by centrifugation (8,000 × *g* at 4°C for 1 h), and the supernatants were filtrated by a 0.22-μm filter (Choice filter, Thermo Scientific). To obtain the purified enzymes, the samples were then applied to a 5-mL HisTrap HP column (GE Healthcare). After washing unbound proteins (20 mM imidazole), the target protein was collected by 250 mM imidazole. The concentration of protein was determined using a protein assay kit (BCA Protein Assay Kit; Genstar, Beijing, China) with BSA as the standard, and the purity of each protein was checked by SDS-PAGE analysis.

PET depolymerization assay to screen variants using gf-PET films

To evaluate the variants' activities, the amorphous gf-PET film (Goodfellow, 1 mm thickness, 6 mm diameter, roughly 67 mg) was soaked in 1,980 mL of 50 mM Tris-HCl buffer (pH 8.5) with 20 μL of purified enzymes at different temperatures (40°C–70°C) for 24 h. The reactions were terminated by heat treatment (100°C, 10 min). The supernatant obtained by centrifugation (8,000 × *g*, 5 min) was then analyzed by Agilent 1200 and Ultimate 3000 UHPLC systems equipped with a Welch Ultimate XBC18 column (4.6 × 250 mm, 5 μm, Welch Materials, Shanghai, China) to quantify PET monomers released from PET depolymerization. The mobile phase was 40% methanol with 0.12% acetic acid (pH 2.5) at a flow rate of 0.7 mL min^{-1}, and the effluent was monitored at a wavelength of 240 nm.

Variant screening for effective hydrolysis of BHET into TPA

When using BHET as the reaction substrate, appropriate amounts of crude wild-type BsEst and each colony of the variant were incubated with 5 mM BHET in a buffer containing 50 mM Tris-HCl (pH 7.5) at different temperatures (30°C–60°C), and samples were taken regularly. The supernatant was obtained by centrifugation (8,000 × *g*, 5 min) and analyzed by HPLC.

Please cite this article in press as: Li et al., Integrating graph learning and meta-learning to enhance PET hydrolase activity at elevated temperatures, Cell Reports Physical Science (2025), https://doi.org/10.1016/j.xcrp.2025.103037

Depolymerization of untreated post-consumed PET products by a two-enzyme system

PET films for the experiment were obtained from commercial suppliers and cut into 6-mm-diameter pieces (12–23 mg). The commercial post-consumed PET (pc-PET) was treated by a two-enzyme system, combining FAST-PETase/DepoPETase variants with BsEst in 1,960 μL of 50 mM Tris-HCl (pH 8.5). The reaction mixture was incubated at 60°C, and the fresh enzyme solution was supplemented every 24 h for depolymerization. The reactions were then terminated by heat treatment (100°C, 10 min). The supernatant obtained by centrifugation (8,000 × *g*, 5 min) was then analyzed by HPLC.

SEM

The morphology of PET films before and after enzyme exposure was examined following a previously reported procedure.[8]

Water contact angle

The water contact angle of PET films before and after enzyme exposure was examined following a previously reported procedure.[33]

Cold-start variant selection

Screening the optimal pose and candidate substitutions

The SMILES format file of substrate 2-HE(MHET)_4 and protein embedding from the ESM2 model (esm2_t36_3B_UR50D)[47] are used as input to perform molecular docking via DiffDock. The parameters are as follows: inference_steps = 20, samples_per_complex = 40, batch_size = 10, actual_steps = 18 and no_final_step_noise. Based on confidence score, the top-1 prediction ligand pose is selected as the optimal binding structure.

To identify potential sites for mutation, considering the spatial relationship between the ligand and the enzyme, a specific distance is established for the unknown active site, such as the Euclidean distance of 10 Å. The range of amino acids from the substrate molecule or the top 50 or top 70 amino acids closest to the active site of the enzyme, based on C_α distance, are then selected as candidate sites for mutation, excluding the active sites. Subsequently, the selection of candidate sites and the type of amino acid will be determined through a GNN.

GVP-based GNN

We employed GVP-GNN,[48] leveraging the advantages of GNNs and CNNs. GVP-GNN replaces dense layers with GVPs within a GNN, predicting the distribution of amino acid types on the protein sequence based on the residue's chemical and topological environment by exchanging information with neighboring residues.

In the GVP-GNN framework, a protein structure is represented as a proximity graph with vertices and edges extracted from 3D backbone atoms. Both vertices and edges contain scalar and vector features. In detail, each vertex that expresses a residue's attributes has embedding with the following features.

(1) Scalar features$\{\sin,\cos\}\circ\{\phi,\psi,\omega\}$, where ϕ, ψ, and ω are the dihedral angles computed from C_{i-1}, N_i, C_{α_i}, C_i, and N_{i+1}.
(2) The forward and reverse unit vectors in directions of $C_{\alpha_{i+1}} - C_{\alpha_i}$ and $C_{\alpha_{i-1}} - C_{\alpha_i}$, respectively.
(3) The unit vector in the imputed direction of $C_{\beta_i} - C_{\alpha_i}$.
(4) A one-hot representation of amino acid identity.

Each edge that expresses the relationship between two residues has embedding with the following features.

(1) The unit vector in the direction of $C_{\alpha_j} - C_{\alpha_i}$.
(2) The encoding of the distance $\|C\_(\alpha\_j) - C\_(\alpha\_i)\|_2$ in terms of Gaussian radial basis functions.
(3) A sinusoidal encoding of $j - i$, which represents distance along the backbone.

Additionally, we introduce the isoelectric point (pI) as a vital amino acid attribute. All features are concatenated and passed to a GVP layer to obtain an initial input graph. Subsequently, GVP-GNN utilizes message passing to update each vertex's embeddings based on messages from neighboring vertices and edges. The output graph, derived from message-passing layers, provides a refined vertex representation, which is further processed through a GVP layer and a SoftMax layer for multi-class classification, resulting in the probability distribution of the 20 standard amino acids across all positions.

To further balance model generalization and biological interpretability, we restricted our training and evaluation to well-curated enzyme domains. Specifically, we adopted the preprocessed CATH 4.2 dataset released by Townshend et al.[49] This dataset has been widely used for benchmarking structure-aware learning models, enabling direct comparison with prior studies.

For each enzyme, residues were mapped to their corresponding 3D coordinates, and the catalytic center was defined based on known active-site annotations or bound ligand positions when available. Residues within a defined spatial radius (typically 8–12 Å) from the active-site centroid were retained as the mutationally tractable region. This heuristic reduces computational complexity while prioritizing positions most likely to modulate catalytic efficiency or substrate recognition. To ensure robustness, residues in flexible loops adjacent to the active site were also considered when sufficient structural confidence scores were available. We acknowledge that our approach does not capture all potential long-range effects, and the relatively modest improvement observed for the BsEst scaffold may partly reflect scaffold-specific constraints and dataset bias.

We retrained the GVP-GNN with the CATH data, following the structural split methodology in previous work: learning rate = 0.001, betas = (0.9, 0.999); training 40 epochs, save best model. The training dataset for our GVP-GNN model was derived from CATH v.4.2, the widely used structural classification database for protein domains. To ensure consistency with previous work, we followed the same dataset preparation and train-test splitting strategy used in the original GVP paper, including sequence identity thresholds and structural redundancy filtering to reduce overfitting.

During model training, we used the Adam optimizer with a mini-batch size of 3,000 amino acid residues. The model was trained to minimize the cross-entropy loss over the 20 standard amino acid classes at each residue position. We explored the hyperparameter space through 50 independent trials with the following settings. Learning rate: log-sampled between 1e−5 and 5e−4; dropout rate: uniformly sampled between 0.05 and

0.5; epochs: 40 maximum, with early stopping if validation loss plateaued (<1% improvement across 5 epochs). The optimal configuration was identified as: learning rate: 0.001; dropout rate: 0.1; β_1 and β_2: default Adam values.

This setup achieved optimal generalization on a held-out validation subset of the CATH 4.2 dataset.

According to the GVP results, there could be a potential variant when the difference between the original type of amino acid and the predicted type with the highest probability of amino acid takes place in one site. Because of the impossibility of validating all variants, the variants that are around the active site or ligand can be selected as experimental candidates. FAST-PETase and DepoPETase share identical active sites (S134/D180/H211). Despite having the same top-50 distance closed candidates, they exhibit distinct GVP results. Consequently, for both enzymes, we selected the ten closest substitutions to the active site for experimentation. In the case of BsEst, which features top 70 substitutions closest to the active site (S189/E310/H399), we chose the 20 substitutions nearest to the active site based on GVP results. We note that the distance-to-active-site threshold was chosen as a pragmatic heuristic to focus mutational exploration on residues most likely to impact activity. While this choice inevitably constrains the search space and does not guarantee a linear relationship with functional changes, it provided a cost-effective strategy under experimental capacity limits.

Fine-tuning the ESM2 model with homologous protein data

Based on the known PETase and BHETase, the hhblits tool helps grab 28,397 sequences with parameters $n = 3$, $B = 100{,}000$, and cpu = 16. According to the principle of Devlin et al.,[50] we fine-tune the ESM2 model as follows.

(1) 10% of the amino acids are masked.
(2) In 80% of the cases, the masked amino acids are replaced by <mask>.
(3) In 10% of the cases, the masked amino acids are replaced by a random amino acid (different) from the one they replace.
(4)In the remaining 10% of cases, the masked amino acids are left as is. Maximum steps = 5,000, learning rate is 1e−4, optimization is a factor 8 V100, 5 days.

Following the fine-tuning of the ESM2 model, we employ it for the design of BHETase (BsEst). Once the embedding of BsEst is obtained, we utilize both the learn2learn model and the Gaussian process regression (GPR) model to predict variants for the next round.

learn2learn model

In this study, we employed the MAML algorithm[51] to discern universal catalytic activity knowledge across various temperature conditions and then transfer this knowledge for robust predictions under new conditions. During the initial experimental round, catalytic activity measurements were obtained at multiple temperatures, and the data were subsequently partitioned into support data and query data.

Recognizing the challenge of identifying substitutions with significant catalytic efficiency based on limited samples, our focus shifted toward achieving enhanced performance using machine-learning techniques, specifically those falling under the umbrella of "learning to rank." This approach encompasses listwise methods, where a ranking function is learned by treating individual lists as instances and minimizing a loss function defined on the predicted list against the ground-truth list. One such listwise method employed in our study is ListMLE, which formalizes learning to rank as a problem of minimizing the likelihood loss function. This method has demonstrated superior properties and yielded improved experimental results on benchmark datasets. The loss function is defined as follows:

$$L(\{y\},\{s\}) = -\log\left(P(\pi_y|s)\right),$$

where $P(\pi_y|s)$ is the Plackett-Luce probability of a permutation π_y conditioned on scores s. Here, π_y represents a permutation of items ordered by the relevance labels y where ties are broken randomly. In the MAML network, we used three multi-layer perceptrons.

GPR model

The GPRs[52] are defined over our parameter space, the set of all combinations of embedding that obtained from fine-tuned ESM2 model, as linear combinations of kernel functions. The kernel functions depend on hyperparameters, most often correlation length scales and signal variances. The GPR model can be trained and conditioned on the experimental data, resulting in an acquisition function that can be passed to an optimizer to find the optimal next measurement outcome.

Defining a GPR model from data $D = \{(x_1,y_1),\ldots,(x_N,y_N)\}$, where $x_i \in R^n$, $n = 1{,}280$, and $y_i = f(x_i) + \epsilon(x_i)$, is accomplished in a Bayesian framework by placing a Gaussian probability density function:

$$p(f) = \frac{1}{\sqrt{(2\pi)^N|\mathbf{K}|}}\exp\left[-\frac{1}{2}(\mathbf{f}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{K}^{-1}(\mathbf{f}-\boldsymbol{\mu})\right],$$

called the prior, over a function space, and condition it on the data. $\boldsymbol{\mu} = [\mu(\mathbf{x_1}),\ldots,\mu(x_N)]^{\mathbf{T}}$ is the mean of the prior Gaussian probability density function, $\mathbf{f} = [f(\mathbf{x_1}),\ldots,f(x_N)]^{\mathbf{T}}$; $K_{ij} = k(\phi,x_i,x_j)$ is the covariance function or kernel, where $\phi$is a set of hyperparameters, commonly length scales l and signal variance σ_s^2. For this study, we used a combination kernel of Dot-product and White. The Dot-product kernel is given by

$$k(x_i,x_j) = \sigma_s^2 + x_i\cdot x_j, \text{where } \sigma_s^2 = 1.0.$$

With $\sigma_s^2 = 1$, the kernel function exhibits a bias term at the origin, making it a nonhomogeneous linear kernel. The nonhomogeneous linear kernel can handle linear relationships beyond those passing through the origin, adding flexibility to the model.

The White kernel is given by

$$k(x_i,x_j) = \begin{cases} 1.0, \text{if } x_i = x_j, \\ 0, \text{if } x_i \neq x_j . \end{cases}$$

Meta-learning framework and model architecture

In the meta-learning phase (round II), we implemented a model-agnostic framework using the learn2learn library to enable rapid adaptation from limited training data.[53] Each training episode consisted of randomly sampled enzyme-specific tasks that were split into support and query subsets (typically at a 3:1 ratio). Meta-batches were constructed by combining multiple episodes to ensure balanced gradient updates and mitigate task-level bias.

The underlying base model was a two-layer fully connected neural network that mapped 1,280-dimensional ESM2 embeddings to a scalar stability score. The model was optimized using the ListMLE loss function, which directly optimizes ranking consistency between predicted and experimentally observed stability, thereby emphasizing the identification of beneficial variants rather than minimizing mean-square error as in conventional regression. Model parameters were updated using meta-gradients derived across tasks through stochastic gradient descent with an adaptive learning rate (1e−4).

To improve generalization, the feature encoder weights from ESM2 were frozen during meta-training, while task-specific adaptation occurred through the final regression layers. Each meta-epoch comprised 50–100 episodes, and convergence was typically achieved after 30–40 epochs. The Gaussian process baseline was trained independently on the same embedding features using a Dot-product + White kernel for computational efficiency in high-dimensional space (n = 1,280).

Collectively, this framework differs from conventional regression pipelines by enabling task-level learning dynamics—that is, the model learns to rank variants effectively across heterogeneous sequence-function relationships and to generalize from sparse supervision. The design allows the identification of promising variants even in low-data regimes where standard regression or evolutionary models such as EvMutation may underperform.

Quantitative assessment of the efficiency of two-enzyme systems by imaging technique

Imaging techniques were employed to establish a quantitative assessment of the efficiency of the two-enzyme system. Taking the example of a two-enzyme system coupled with DepoPETase, we initially obtained SEM images of PET film processed through the system for 48 h. These included the control image (PET film), the DepoPETase (single-enzyme system) catalyzed image, the DepoPETase/BsEst (two-enzyme system) catalyzed image, and the Depo-S181Y/BsEst-Y109R (two-enzyme system) catalyzed image (Figures 5G and 5H). After removal of watermarks, we calculated the difference between the experimental images and control images, respectively. In the control group, the SEM image has a smooth surface with minimal brighter pixel dots. As the PET film underwent erosion by single-/two-enzyme systems, the PET film surface became rougher, evident in the images by the emergence of bright pixels. The intensity of the pixel brightness correlated with the roughness of the membrane surface and the catalytic capacity under identical reaction conditions. Subsequently, the SNR was calculated using the formula

$$\mathrm{SNR} = \frac{\mu_{\mathrm{sig}}}{\sigma_{\mathrm{sig}}},$$

which is defined as the ratio of the average image signal value μ_{sig} to the standard deviation of the image signal σ_{sig}. A higher SNR value indicates superior image quality, reflecting a higher mean value of the image signal, more bright spots, and a rougher surface of the PET film. Simultaneously, a smaller variance in the image signal signifies less fluctuation in signal data, indicating greater consistency in image signal values.

Molecular docking and MD simulation

Given that the crystal structure of PETase is known (PDB: 6EQE), we used it as a reliable template for predicting the structures of the FAST-PETase and DepoPETase variants with AlphaFold 2, which resulted in high prediction accuracy.[54] Representative structure was then used for further molecular docking. A previously reported model substrate including 4PET and BHET was used.[33,55] Molecular docking was performed using AutoDock Tools 1.5.6.[56] The predicted catalytic residues were used to define the binding pocket, and clustering analysis of the output results was conducted by AutoDock using criteria such as energy minimization and cluster size. The energetically favorable poses of ligands binding to the targeted sites of enzymes were extracted and analyzed. After yielding the final docked binding mode, MD simulation and analysis were performed with the GROMACS 2016 simulation package with the GROMOS96 (54a7) force field. Docking conformations and topology files of substrates 4PET, BHET, and intermediates BHET, MHET, EG, and TPA were obtained from the ATB website (http://atb.uq.edu.au/index.py). The docking conformation and topology files of FAST-PETase, FAST-Y61Q-PETase, DepoPETase, and Depo-S181Y-PETase maintained the docked conformation with 4PET in the simulated system, and the intermediates BHET, MHET, EG, and TPA molecules were randomly inserted into the system. BsEst, Bs-Y109R, and Bs-L465V maintained the docked conformation with BHET in the simulated system. A total of 42 simulated reactions at 303 K and 333 K after docking with substrates were performed in the system for the two types of proteins PETase and BsEst, respectively. The proteins, substrates, and intermediates were first placed in a cubic box with a minimum distance of 12 Å from the edge of the box to the proteins, and the box was next filled with water molecular model simple point charge extended (SPCE). To equilibrate the system, Na^+ and Cl^- were added for net charge neutralization. To avoid unfavorable interactions, energy minimization was performed using the steepest descent method prior to MD simulations, followed by 100 ps canonical ensemble treatments (NVT) at 303 K and 333 K, then 100 ps constant-pressure, constant-temperature treatments (NPT), and finally running the simulations (100 ns at 303 K/333 K, 1 bar, and 2 fs time step). To avoid experimental chance introduced by a single simulation, we used three simulations with different starting atomic velocities as parallel controls. During the simulations, the energy, coordinates, and velocity were recorded at 0.5 ns intervals, and the trajectories were visualized and analyzed with PyMOL 2.5.2 and VMD 1.9.3; all analyses were finally calculated using the GROMACS simulation packaging tool.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Xiujuan Li (lixiujuan@njnu.edu.cn).

Materials availability

This study did not generate new, unique reagents.

Data and code availability

- The data reported in this paper and the supplemental information will be shared by the lead contact upon request.
- Source code in this work is available at https://github.com/yibos8186-source/Integrating-GVP-GNN-with-Meta-learning.git.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (2024YFA0917800), the National Natural Science Foundation of China (22478199), the Basic Research Program of Jiangsu (BK20250141), and the Jiangsu Basic Research Center for Synthetic Biology (BK20233003). We would like to thank the Hangzhou Zhejiang Lab and Hangzhou LUCA Intelligent Technology for their invaluable contributions to data collection and model training, which significantly enhanced the outcomes of this work.

AUTHOR CONTRIBUTIONS

The project was coordinated by X.L., H.H., and S.L.; X.L., A.L., and Y.S. jointly conceptualized the project and designed the research experiments; J.H., J.T., and Z.L. executed all the ML studies; A.L. performed experimental work; Y.S. performed MD simulations; and A.L., H.Z., and Y.S. analyzed the results. All authors contributed to writing and revising the manuscript, and all authors have read and agreed to the published version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrp.2025.103037.

Received: July 3, 2025
Revised: October 30, 2025
Accepted: November 25, 2025

REFERENCES

1. Harnkarnsujarit, N., Wongphan, P., Chatkitanan, T., Laorenza, Y., and Srisa, A. (2021). Bioplastic for Sustainable Food Packaging. In Sustainable Food Processing and Engineering Challenges (Elsevier), pp. 203–277. https://doi.org/10.1016/B978-0-12-822714-5.00007-3.
2. Rosenboom, J.-G., Langer, R., and Traverso, G. (2022). Bioplastics for a circular economy. Nat. Rev. Mater. *7*, 117–137.
3. Han, J., Guo, Y., Wang, H., Zhang, K., and Yang, D. (2021). Sustainable Bioplastic Made from Biomass DNA and Ionomers. J. Am. Chem. Soc. *143*, 19486–19497.
4. Yan, N. (2022). Recycling plastic using a hybrid process. Science *378*, 132–133.
5. Walker, S., and Rothman, R. (2020). Life cycle assessment of bio-based and fossil-based plastic: A review. J. Clean. Prod. *261*, 121158.
6. Zimmermann, L., Dombrowski, A., Völker, C., and Wagner, M. (2020). Are bioplastics and plant-based materials safer than conventional plastics? In vitro toxicity and chemical composition. Environ. Int. *145*, 106066.
7. Lu, H., Diaz, D.J., Czarnecki, N.J., Zhu, C., Kim, W., Shroff, R., Acosta, D.J., Alexander, B.R., Cole, H.O., Zhang, Y., et al. (2022). Machine learning-aided engineering of hydrolases for PET depolymerization. Nature *604*, 662–667.
8. Cui, Y., Chen, Y., Liu, X., Dong, S., Tian, Y., Qiao, Y., Mitra, R., Han, J., Li, C., Han, X., et al. (2021). Computational redesign of a PETase for plastic biodegradation under ambient condition by the GRAPE strategy. ACS Catal. *11*, 1340–1350.
9. Bell, E.L., Smithson, R., Kilbride, S., Foster, J., Hardy, F.J., Ramachandran, S., Tedstone, A.A., Haigh, S.J., Garforth, A.A., Day, P.J.R., et al. (2022). Directed evolution of an efficient and thermostable PET depolymerase. Nat. Catal. *5*, 673–681.
10. Guo, B., Lopez-Lorenzo, X., Fang, Y., Bäckström, E., Capezza, A.J., Vanga, S.R., Furó, I., Hakkarainen, M., and Syrén, P.O. (2023). Fast Depolymerization of PET Bottle Mediated by Microwave Pre-Treatment and An Engineered PETase. ChemSusChem *16*, e202300742.
11. Shi, L., Liu, P., Tan, Z., Zhao, W., Gao, J., Gu, Q., Ma, H., Liu, H., and Zhu, L. (2023). Complete depolymerization of PET wastes by an evolved PET hydrolase from directed evolution. Angew. Chem. Int. Ed. *62*, e202218390.
12. Tournier, V., Topham, C.M., Gilles, A., David, B., Folgoas, C., Moya-Leclair, E., Kamionka, E., Desrousseaux, M.L., Texier, H., Gavalda, S., et al. (2020). An engineered PET depolymerase to break down and recycle plastic bottles. Nature *580*, 216–219.
13. Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., Toyohara, K., Miyamoto, K., Kimura, Y., and Oda, K. (2016). A bacterium that degrades and assimilates poly(ethylene terephthalate). Science *351*, 1196–1199.
14. Eberl, A., Heumann, S., Brückner, T., Araujo, R., Cavaco-Paulo, A., Kaufmann, F., Kroutil, W., and Guebitz, G.M. (2009). Enzymatic surface hydrolysis of poly(ethylene terephthalate) and bis(benzoyloxyethyl) terephthalate by lipase and cutinase in the presence of surface active molecules. J. Biotechnol. *143*, 207–212.
15. Sulaiman, S., You, D.-J., Kanaya, E., Koga, Y., and Kanaya, S. (2014). Crystal structure and thermodynamic and kinetic stability of metagenome-derived LC-cutinase. Biochemistry *53*, 1858–1869.
16. Herrero Acero, E., Ribitsch, D., Steinkellner, G., Gruber, K., Greimel, K., Eiteljoerg, I., Trotscha, E., Wei, R., Zimmermann, W., Zinn, M., et al. (2011). Enzymatic surface hydrolysis of PET: Effect of structural diversity on kinetic properties of cutinases from *Thermobifida*. Macromol. *44*, 4632–4640.
17. Zhang, Z., Huang, S., Cai, D., Shao, C., Zhang, C., Zhou, J., Cui, Z., He, T., Chen, C., Chen, B., and Tan, T. (2022). Depolymerization of post-consumer PET bottles with engineered cutinase 1 from Thermobifida cellulosilytica. Green Chem. *24*, 5998–6007.
18. Ribitsch, D., Acero, E.H., Greimel, K., Eiteljoerg, I., Trotscha, E., Freddi, G., Schwab, H., and Guebitz, G.M. (2012). Characterization of a new cutinase from *Thermobifida alba* for PET-surface hydrolysis. Biocatal. Biotransform. *30*, 2–9.
19. Liu, Y., Liu, Z., Guo, Z., Yan, T., Jin, C., and Wu, J. (2022). Enhancement of the degradation capacity of IsPETase for PET plastic degradation by protein engineering. Sci. Total Environ. *834*, 154947.
20. Shi, L., Liu, P., Tan, Z., Zhao, W., Gao, J., Gu, Q., Ma, H., Liu, H., and Zhu, L. (2023). Complete depolymerization of PET waste by an evolved PET hydrolase from directed evolution. Angew. Chem. Int. Ed. *62*, e202218390.
21. Son, H.F., Cho, I.J., Joo, S., Seo, H., Sagong, H.-Y., Choi, S.Y., Lee, S.Y., and Kim, K.-J. (2019). Rational protein engineering of thermo-stable PETase from *Ideonella sakaiensis* for highly efficient PET degradation. ACS Catal. *9*, 3519–3526.
22. Cui, Y., Sun, J., and Wu, B. (2022). Computational enzyme redesign: large jumps in function. Trends Chem. *4*, 409–419.

23. Torrisi, M., Pollastri, G., and Le, Q. (2020). Deep learning methods in protein structure prediction. Comput. Struct. Biotechnol. J. *18*, 1301–1310.

24. Meier, J., Rao, R., Verkuil, R., Liu, J., Sercu, T., and Rives, A. (2021). Language models enable zero-shot prediction of the effects of mutations on protein function. Preprint at bioRxiv. https://doi.org/10.1101/2021.07.09.450648.

25. Hu, R., Fu, L., Chen, Y., Chen, J., Qiao, Y., and Si, T. (2023). Protein engineering via Bayesian optimization-guided evolutionary algorithm and robotic experiments. Brief. Bioinform. *24*, bbac570.

26. Steier, A., Muñiz, A., Neale, D., and Lahann, J. (2019). Emerging Trends in Information-Driven Engineering of Complex Biological Systems. Adv. Mater. *31*, 1806898.

27. Cui, Y., Chen, Y., Sun, J., Zhu, T., Pang, H., Li, C., Geng, W.C., and Wu, B. (2024). Computational redesign of a hydrolase for nearly complete PET depolymerization at industrially relevant high-solids loading. Nat. Commun. *15*, 1417.

28. Ding, Z., Xu, G., Miao, R., Wu, N., Zhang, W., Yao, B., Guan, F., Huang, H., and Tian, J. (2023). Rational redesign of thermophilic PET hydrolase LCCICCG to enhance hydrolysis of high crystallinity polyethylene terephthalates. J. Hazard. Mater. *453*, 131386.

29. Pörtner, H.O., Bennett, A., Bozinovic, F., Clarke, A., Lardies, M., Lucassen, M., Pelster, B., Schiemer, F., and Stillman, J. (2006). Trade-Offs in Thermal Adaptation: The Need for a Molecular to Ecological Integration. Physiol. Biochem. Zool. *79*, 295–313.

30. Kaltenbach, M., Emond, S., Hollfelder, F., and Tokuriki, N. (2016). Functional Trade-Offs in Promiscuous Enzymes Cannot Be Explained by Intrinsic Mutational Robustness of the Native Activity. PLoS Genet. *12*, e1006305.

31. Stigter, D., and Dill, K.A. (1990). Charge effects on folded and unfolded proteins. Biochemistry *29*, 1262–1271.

32. Bjellqvist, B., Hughes, G.J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.C., Frutiger, S., and Hochstrasser, D. (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis *14*, 1023–1031.

33. Li, A., Sheng, Y., Cui, H., Wang, M., Wu, L., Song, Y., Yang, R., Li, X., and Huang, H. (2023). Discovery and mechanism-guided engineering of BHET hydrolases for improved PET recycling and upcycling. Nat. Commun. *14*, 4169.

34. Hopf, T.A., Ingraham, J.B., Poelwijk, F.J., Schärfe, C.P.I., Springer, M., Sander, C., and Marks, D.S. (2017). Mutation effects predicted from sequence co-variation. Nat. Biotechnol. *35*, 128–135.

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

36. Wei, R., von Haugwitz, G., Pfaff, L., Mican, J., Badenhorst, C.P.S., Liu, W., Weber, G., Austin, H.P., Bednar, D., Damborsky, J., and Bornscheuer, U.T. (2022). Mechanism-based design of efficient PET hydrolases. ACS Catal. *12*, 3382–3396.

37. Chen, Z., Duan, R., Xiao, Y., Wei, Y., Zhang, H., Sun, X., Wang, S., Cheng, Y., Wang, X., Tong, S., et al. (2022). Biodegradation of highly crystallized poly(ethylene terephthalate) through cell surface codisplay of bacterial PETase and hydrophobin. Nat. Commun. *13*, 7138.

38. Dewi, S., Horbach, T., Neumann, T., Jaeger, K., Streit, W., and Hartmann, L. (2025). Evaluation of pretreatment strategies and combinations of enzymes for degradation of polyethylene terephthalate. J. Polym. Sci. *63*, 2770–2781.

39. Thomsen, T.B., Almdal, K., and Meyer, A.S. (2023). Significance of poly(-ethylene terephthalate) (PET) substrate crystallinity on enzymatic degradation. Nat. Biotechnol. *78*, 162–172.

40. Brizendine, R.K., Erickson, E., Haugen, S.J., Ramirez, K.J., Miscall, J., Salvachúa, D., Pickford, A.R., Sobkowicz, M.J., McGeehan, J.E., and Beckham, G.T. (2022). Particle size reduction of poly(ethylene terephthalate) increases the rate of enzymatic depolymerization but does not increase the overall conversion extent. ACS Sustainable Chem. Eng. *10*, 9131–9140.

41. Cui, H., Vedder, M., Schwaneberg, U., and Davari, M.D. (2022). Using molecular simulation to guide protein engineering for biocatalysis in organic solvents. Methods Mol. Biol. *2397*, 179–202.

42. Tan, K.P., Varadarajan, R., and Madhusudhan, M.S. (2011). DEPTH: a web server to compute depth and predict small-molecule binding cavities in proteins. Nucleic Acids Res. *39*, W242–W248.

43. Lin, Q., Zhan, Q., Li, R., Liao, S., Ren, J., Peng, F., and Li, L. (2021). Solvent effect on xylose-to-furfural reaction in biphasic systems: combined experiments with theoretical calculations. Green Chem. *23*, 8510–8518.

44. Fan, F., Zheng, Y., Zhang, Y., Zheng, H., Zhong, J., and Cao, Z. (2019). A comprehensive understanding of enzymatic degradation of the G-type nerve agent by phosphotriesterase: revised role of water molecules and rate-limiting product release. ACS Catal. *9*, 7038–7051.

45. Feng, S., Yue, Y., Zheng, M., Li, Y., Zhang, Q., and Wang, W. (2021). *Is* PETase- and *Is* MHETase-Catalyzed Cascade Degradation Mechanism toward Polyethylene Terephthalate. ACS Sustainable Chem. Eng. *9*, 9823–9832.

46. Mrigwani, A., Thakur, B., and Guptasarma, P. (2022). Conversion of polyethylene terephthalate into pure terephthalic acid through synergy between a solid-degrading cutinase and a reaction intermediate-hydrolysing carboxylesterase. Green Chem. *24*, 6707–6719.

47. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., and Rives, A. (2022). Language models of protein sequences at the scale of evolution enable accurate structure prediction. Preprint at bioRxiv. https://doi.org/10.1101/2022.07.20.500902.

48. Jing, B., Eismann, S., Suriana, P., Townshend, R.J.L., and Dror, R. (2021). Learning from Protein Structure with Geometric Vector Perceptrons. Preprint at arXiv. https://doi.org/10.48550/arXiv.2009.01411.

49. Townshend, R.J.L., Vögele, M., Suriana, P., Derry, A., Powers, A., Laloudakis, Y., Balachandar, S., Jing, B., Anderson, B., Eismann, S., et al. (2022). ATOM3D: tasks on molecules in three dimensions. Preprint at arXiv. https://doi.org/10.48550/arXiv.2012.04035.

50. Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Preprint at arXiv. https://doi.org/10.48550/arXiv.1810.04805.

51. Finn, C., Abbeel, P., and Levine, S. (2017). Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. Preprint at arXiv. https://doi.org/10.48550/arXiv.1703.03400.

52. Rasmussen, C.E., and Williams, C.K.I. (2008). Gaussian Processes for Machine Learning (MIT Press).

53. Arnold, S.M.R., Mahajan, P., Datta, D., Bunner, I., and Zarkias, K.S. (2020). learn2learn: a library for meta-learning research. Preprint at arXiv. https://doi.org/10.48550/arXiv.2008.12284.

54. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589.

55. Han, X., Liu, W., Huang, J.W., Ma, J., Zheng, Y., Ko, T.P., Xu, L., Cheng, Y.S., Chen, C.C., and Guo, R.T. (2017). Structural insight into catalytic mechanism of PET hydrolase. Nat. Commun. *8*, 2106.

56. Goodsell, D.S., Sanner, M.F., Olson, A.J., and Forli, S. (2021). The AutoDock suite at 30. Protein Sci. *30*, 31–43.