Proceedings of the 4th International Conference on Biological Engineering and Medical Science
DOI: 10.54254/2753-8818/67/2024.18882

Simulation Assisted Improvement of Plastic Degradation Enzyme PETase based Machine Learning Tools

Ziming Tian

Shanghai American School Puxi Campus, Shanghai, China

diegotian2007@hotmail.com

Abstract. Polyethylene terephthalate (PET) plastic is one of the most widely used plastic primarily due to its flexibility, endurance, and low cost. However, the plastic's one-time use nature and long degradation time have led to massive waste accumulation, damaging our ecosystem, health, and biodiversity. While previous degradation methods are ineffective due to their high cost and low efficiency, the discovery of two enzymes PETase and MHETase in the bacteria *Ideonella sakaiensis* to degrade PET and mono(2-hydroxyethyl), a reaction intermediate in PET degradation, respectively, sparked the idea of a sustainable approach to degradation. Ever since, many approaches, including directed evolution, rational protein engineering, and computational redesign strategies, have optimized PETase in terms of its thermostability, catalytic activity, and more. This study proposes the incorporation of newly developed machine learning-based computational tools, including MutCompute, AlphaFold, and DiffDock, into a holistic protein engineering process to predict optimal PETase mutations. Here, in-silico experiments using machine learning tools as well as molecular dynamics simulation and interactions analysis screened for large amounts of PETase mutants in a time and cost-saving manner. Degradation assay coupled with mass analysis and high-performance liquid chromatography techniques then experimentally characterized PETase and its chosen mutants; thus, further screening found the most viable PETase mutant. Using various strategies, the project directly tackles one of the major global issues – sustainability – by bio-recycling PET. The research also aims to pave the way for introducing a new, imitable process for the more effective and resource-efficient engineering of all proteins.

Keywords: Machine learning, protein engineering, sustainability, PETase, plastic degradation.

1. Introduction

Plastic pollution is a pressing issue that society has to deal with. Plastic produced from human use is often released into the environment, which has historically contributed to the formation of floating islands in the ocean (the infamous Great Pacific Garbage Patch, for example) that has dealt serious environmental damages to marine species [1]. Moreover, plastic appears in the form of microplastic present within animals and plants that are consumed by humans, revealing the significant health implications of this issue [2]. In short, plastic pollution poses a threat to the environment, human health, and the entire ecosystem. All of this roots from the fact that plastic has slowly become the irreplaceable material for societal development nowadays. Plastic widely refers to a range of materials built from polymers of high molecular weight [3] that are desired due to their strength, flexibility, durability, and lightness [1]. Currently on the market is a wide variety of plastics that serve diverse functions.

DOI: 10.54254/2753-8818/67/2024.18882

One of the most prominent plastics currently is polyethylene terephthalate (often abbreviated as PET), a type of polyester plastic polymerized from ethylene glycol (EG) and terephthalic acid (TPA) [4]. In fact, this type of plastic has many attractive qualities such as strength, heat and chemical resistance, as well as being nonreactive when contacting food and water [5]. PET is also a thermoplastic, which means that it can be molded easily [5]. As a result, PET is extensively used for plastic water bottles, food packaging, clothing, and fishing nets [4]. As of 2021, more than 82 million metric tons of PET were reported to be produced each year with the market for PET products growing continuously [6]. Its production is especially increased due to the one-time use nature of many PET products. PET is petroleum-based [6] and does not form naturally, and the natural degradation of the durable polymer by microorganisms in the environment is estimated to take hundreds of years [5]. All of which reaffirms PET's problematic role within our ecosystem.

To address the problem of PET pollution, ongoing research is focused on PET degradation into smaller, recyclable products [7]. Chemical degradation using alcoholysis holds potential in PET degradation [8], Yet, it is costly and technically challenging. The aromatic components of PET cause it to be chemically inert [7]. PET is relatively resistant to mechanical stress and heat due to its polymer strength, making these two degradation methods also ineffective. Recently, the microorganism *Ideonella sakaiensis* was identified to be an organism that uses PET as its main energy and carbon source. Two enzymes PETase and MHETase were successfully isolated to biologically degrade PET and one of its products, mono(2-hydroxyethyl) terephthalic acid (abbreviated as MHET), into the monomers TPA and EG, respectively (Figure 1a) [4, 7]. These enzymes are naturally evolved and have demonstrated remarkable activity against highly crystalline forms of PET [4]. This opened up the future path for sustainable PET degradation.

Still, further enhancement made to the enzymes is imperative for large-scale, efficient enzymatic degradation of PET that can decrease cost and time required for degradation [5]. The original enzyme *Is*PETase, derived from Ideonella *sakaiensis*, also possesses a low thermostability [9], meaning that it will denature at a relatively low temperature. This characteristic prevents the enzyme from taking advantage of the glass transition temperature of PET (around 60-70° C) where the high thermal energy is able to facilitate the breaking of chemical bonds within the polymer, which is at a very flexible state, for more efficient catabolism [10].

Currently, scientists have indeed achieved mutagenesis of the two enzymes, primarily focusing on PETase, through a variety of methods. Many PETase mutants exhibit higher thermostability and improved performance. For example, HotPETase was engineered using directed evolution [9]; ThermoPETase was produced with rational protein design [11]; and FAST-PETase was developed with a 3D convolutional neural network in MutCompute [10].

Biology-related machine learning simulations, which have advanced significantly in recent years, hold immense potential in aiding the optimization of macromolecules like proteins. They are able to make logical predictions and decisions for humans. This study establishes a holistic, high through-put in-silico protein engineering method based on machine learning tools. We show through its optimization of PETase its effectiveness and potential in being applied to many more proteins in the future.

We used MutCompute to predict potential sites of mutagenesis [12]. For each new PETase sequence, a site was mutated once each for one of the other 19 common amino acids. These sequences were then predicted using AlphaFold to acquire 3D protein structure [13]. DiffDock [14, 15] was employed to dock a PET quadromer onto each mutant PETase, which screened for best binding affinity and docking confidence. Molecular dynamics simulation validated the stability of the complex upon PET docking to the mutant PETase [16]. Interaction analysis using ChimeraX [17] was done on several of the new mutants. Final mutants were chosen for experimental verification. *Escherichia coli* expressed plasmids containing mutant PETase sequences. The cell lysate as well as the purified proteins were incubated with PET granules after evaluation of protein concentration with SDS-PAGE and Bicinchoninic Acid Assay. Mass analysis and high-performance liquid chromatography assay [3] ultimately determined outstanding mutants with improved activity and thermostability, with the most significant mutation causing a 4.5-fold increase in degradation activity at 50° C.

DOI: 10.54254/2753-8818/67/2024.18882

2. Methods

2.1. Computational Methods

2.1.1. Optimal Mutation Position Prediction. Mutation prediction was completed with 3D convolutional neural network MutCompute, a 3D, self-supervised, convolutional neural network trained by learning about the chemical microenvironment near amino acids of more than 19,000 protein structures on the Protein Data Bank (abbreviated as PDB). Upon inputting the PDB ID (6EQE in the case of *Is*PETase) into the online tool (Website link: https://mutcompute.com/), the tool predicted the certainty that each position's wild-type residue is the most optimal as a percentage [12]. Chosen positions had low certainty that the original residue is optimal and we ensured that they are near the active site, in order to make sure that they contribute primarily to catalytic activity.

2.1.2. Mutant Structure Simulation. Selected mutation positions were mutated once for each of the 19 other common amino acids to generate varied amino acid sequences. Its consequent 3D structure was simulated. The deep learning tool AlphaFold 2 was incorporated to fold each new amino acid sequence into a 3D structure by predicting the affected amino acid interactions and levels of protein folding. The AlphaFold 2 model was trained to do so via using information regarding homologous structures and multiple sequence alignments [13]. The code for the AlphaFold model was accessed via the official notebook but had been however adjusted to facilitate batch mutant structure prediction by simply inputting mutation location and amino acids (Link of Google Colab notebook code: https://colab.research.google.com/drive/1TQhWUB-xNMOmonwbEldWLJIBB-5HI8AN?usp=sharing) [18]. PDB files of 3D protein mutants were generated. AlphaFold returned 5 predicted versions of a mutant protein to account for the uncertainty in protein folding.

2.1.3. Ligand Docking Simulation. Using the molecular editor Avogadro, the ligand for the docking simulation was drawn whose molecular geometry was optimized via the built-in "auto optimization tool" [19]. Using the PDB files of predicted mutants and the created ligand, diffusion generative model DiffDock simulated ligand-protein docking by predicting the docked position of PET in regard to the mutant PETase proteins [15]. Within the docking simulation, binding affinity was further optimized via the deep learning software Gnina [20]. Similar to previously, the code had been adjusted to allow batch ligand docking simulations (Link of Google Colab notebook code: https://colab.research.google.com/drive/1gEH5DNu6GVM2cYTdlrn4LN0AWEPHgMnw?usp=sharing). Each docked result returned a variation of potential docking positions, each one having a Gnina-minimized binding affinity and DiffDock-scored docking confidence. Docked results of each mutant were organized into its own series. The series for each mutant was scatter plotted using the Matplotlib library [21], with the x-axis being the docking confidence and the y-axis being the minimized binding affinity. The code was manipulated to only show a specific window of the scatter plot. Optimal mutants were finally identified on the basis that many of them in the same series showed up in the desired window. They also showed exceptional performance in both docking confidence and binding affinity.

2.1.4. Molecular Dynamics Simulation. Proteins underwent molecular dynamics simulation to verify their viability and docking stability. The simulation was run primarily based off the OpenMM toolkit [16] (Link of Google Colab notebook code: https://colab.research.google.com/drive/1IEuae-GeDoeg3-0WICiEDr1jDd2v97HU?usp=sharing). The TIP3P water model was selected and the simulation was run in a 10 Å × 10 Å × 10 Å boundary condition. Ff14SB was utilized as the protein force field while GAFF2 was the ligand force field during the simulation. 0.15 mol/L of NaCL ions neutralized the system. Under a total simulation time of 5 ns, the system underwent the equilibration process in 1000 steps. The final production run occurred with a simulation time of 10 ns at the temperature of 298 K and a pressure of 1 bar in the NPT ensemble. Analysis was done based on returned data, primarily by calculating the root mean square deviation of CA atoms (or, alpha carbons) on the polypeptide chain. Each protein

DOI: 10.54254/2753-8818/67/2024.18882

represented a series with RMSD being plotted with respect to the time during the simulation from 0 to 1000 steps. RMSD of CA atoms throughout the simulation could be compared to the wild-type for assessment of stability [22].

2.1.5. Interactions Visualization. Interactions between the PET quadromer and PETase mutants were observed quickly using the software ChimeraX [17]. Inputting files for each mutant and ligand, ChimeraX facilitated the search for potential hydrogen bonds and π-π stacking interactions [23] between the protein and the ligand. Comparing the visualized results to that of the wild-type, we were able to analyze potentially new interactions introduced by the mutant.

2.2. Experimental Methods

2.2.1. Protein Expression. Gene Construction The genes for the PETase mutants alongside the wild-type enzyme were commercially synthesized by Sangon Biotech and were codon optimized for suitable protein expression in *E. coli* bacteria cells. Each gene was cloned into pET-24b (+) vectors individually.

Transformation The designed vectors were transformed into BL21(DE3) competent *E. coli* cells through chemical competent cell transformation. Cells were initially prepared via incubation within calcium chloride ($CaCl_2$) in order to increase the permeability of the bacterial cell membrane. Cells were thawed on ice, and 20 ng of the constructed plasmid DNA was added to 50 μL of the cells. The mixture was placed on ice for 30 minutes and subsequently heat shocked at 42° C for 40 seconds. 950 μl of S.O.C. Media (without antibiotics) was added to the tube to facilitate transformation. The tube was placed for 60 minutes within a 37° C shaking incubator to be shaken vigorously at 250 rpm. Warm selection plates (specifically, selection plates with Kanamycin for pET-24b (+)) were prepared and heated to 37° C, and 50 μl of the cells were spread onto each plate. Plates were left overnight for incubation at 37° C.

Induced Expression Single colonies of transformed *E. Coli* were resuspended in a liquid culture containing Kanamycin for pET-24b (+). The colony was set again to grow overnight into saturated cell density and it was once again inoculated at a 1:100 dilution into 1L of expression media with Kanamycin for pET-24b (+). The culture was incubated within a 37° C shaking incubator at 200 rpm until the culture reaches 0.4-0.8 for its OD600 value. 0.5 mM of IPTG was then added for induction and proteins were expressed overnight at 30° C. Using centrifugation, cells were harvested from the spent media and resuspended in 1x lysis buffer (300 mM NaCl, 20 mM Tris-HCl, pH 7.5, 1x PMSF, 10 mM β-mercaptoethanol). The cells were disrupted by sonication on ice for 20 minutes with 3 seconds pulses at 7 seconds intervals to obtain the cell lysate. A clearer cell lysate was produced after centrifugation for 10 min at 2,900 g to remove insoluble cell debris, and the precipitate was then ready for downstream analysis.

2.2.2. Protein Purification. The soluble lysate underwent affinity chromatography for protein purification via the application of Ni-NTA resin. 1 mL of Ni-NTA resin was applied to the column initially alongside 4 mL of the cell lysate. The liquid from the purification column was let to flow out under gravity as the bottom lid was opened. The column was washed 5 times with 1 mL non-denaturing washing solution and then underwent 5 rounds of elution for the bound proteins using non-denaturing elution solution. Collected elution solution featured the desired His-tagged PETase protein and was ready for downstream analysis.

2.2.3. Degradation Activity Assessment. Incubation Each mutant under a specific temperature condition had tests in triplicates. 50 mg PET granules were incubated with 2 mL of cell lysate or 1 mL of 1mg/mL purified proteins respectively, to test for degradation activity. All reaction mixtures were incubated for 5 days. For testing with cell lysate, 30, 37, and 50° C reaction conditions were tested for each case while 37 and 50° C were the reaction conditions for the purified protein testing.

DOI: 10.54254/2753-8818/67/2024.18882

Mass Analysis Remaining PET granules were washed with PBS three times and with ultra-pure water for an additional three times. The granules were air dried, the weight was measured, and the percentage degraded was calculated, acting as the simplest assessment to quantify degradation. The average activity and standard deviation were calculated as trials were in triplicates.

High-Performance Liquid Chromatography (HPLC) Assay The solution after the reaction between PETase and PET at 37° C was obtained and the supernatant was extracted with centrifugation for 10 minutes at 15,000 g. Once 0.22-μm filtration was completed, a high-performance liquid chromatography (HPLC) system [3] including a C18 column analyzed 20 μL of the solution that has been filtered. The mobile phase featured methanol with 20 mM of phosphate buffer (pH 2.5) flowing at 0.8 mL per minute with monitoring of washed away solution at 240 nm wavelength. The elution condition was 25 minutes with 25-85% methanol, increasing in a linear gradient. The peak area for the retention time found of compound MHET was calculated from data provided by the machine. The datapoints were averaged and standard deviation was calculated based on the three trials conducted per test case.

2.2.4. Protein Characterization. SDS-PAGE and Coomassie Blue Staining To confirm the abundance of the cell lysate and the purified protein in the reaction mixture, SDS-PAGE assay was initially conducted. The Tris/Glycine SDS Running Buffer (10x), pH 8.3 was applied for the experiment. 10 μL of samples for cell lysate or purified protein was simply loaded into each of the wells of the previously prepared stacking gel. 2 μL of protein ladder was applied for each well on the left and right side. Gel electrophoresis ran at 100 V for 60 minutes. The gel was incubated with 100 mL Coomassie Blue staining solution (0.1% Coomassie Blue in 40% ethanol and 10% acetic acid) for 1 hour on an orbital shaker until the bands become visible. After rinsing one time, a destaining solution (10% ethanol and 7.5% acetic acid) was also applied and left to incubate with the gel on an orbital shaker until the background was nearly clear. The gel was then ready for analysis of visualized results.

BCA Assay Nine diluted BSA samples were prepared for concentrations of 2000, 1500, 1000, 750, 500, 250, 125, 25, 0 μg/mL. The BCA working reagent was prepared to the necessary volume based on the ratio 50:1 for reagent A:B. 0.1 mL of each BSA standard and another case for the unknown samples were pipetted and mixed in a test tube wth 2.0 mL of BCA working reagent. Test tubes were covered and incubated at 37° C in a water bath for 30 minutes. Once the test tube was adjusted back to room temperature, the absorbance of each sample was measured using a spectrophotometer set to 562 nm, with the blank standard's absorbance being subtracted from each. A standard curve plotting each BSA standard's 562 nm absorbance measurement to its concentration was created. Protein concentrations of unknown samples were then determined based off the standard curve.

3. Results and Discussion

3.1. Rationale for Selection of Mutation Site

For our protein engineering efforts, we chose the original *Is*PETase as the base enzyme. Its properties are relatively more known and allow analysis of the characteristics of each mutated position after they are chosen [1].

DOI: 10.54254/2753-8818/67/2024.18882

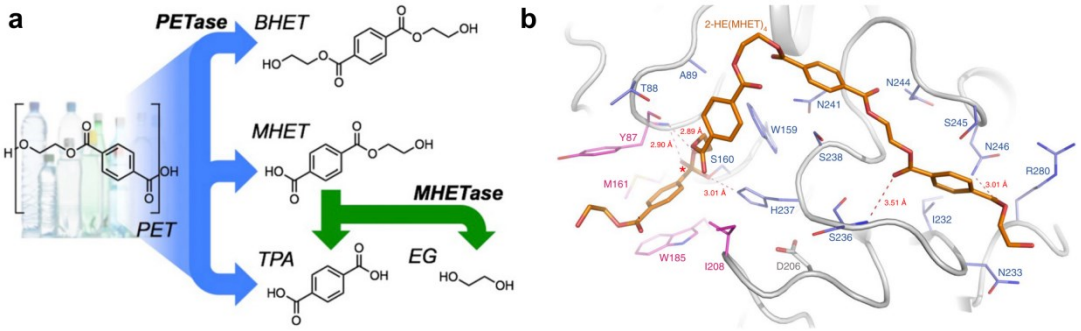

Figure 1. The Degradation Mechanism of IsPETase has been Consolidated by Published Research. a. PET degradation pathway using PETase [4]. PET is degraded into BHET, MHET, and TPA at various amounts. Another complementary enzyme, MHETase, is responsible for further degrading MHET into TPA and EG. b. Residues of IsPETase involved in the degradation of 2-HE(MHET)4 [1]. As the orange ligand 2-HE(MHET)4 is brought into proximity with the PETase enzyme, hydrogen bonds formed are labeled. The responsible residues are also depicted with labeling.

*Is*PETase degrades PET via the same pathway previously mentioned and illustrated by Austin et al. (Figure 1a) It facilitates the breakdown of PET into BHET, MHET, and TPA. While BHET and TPA are produced as products from PET degradation, it is to note though that the majority of the product is created in MHET, a piece of useful information when analyzing the results of degradation using our mutated enzymes [4]. Furthermore, Joo et al. reported in 2018 the main amino acid residues near the active site for *Is*PETase that have been responsible for catabolism of their model of the PET chain, which was 2-HE(MHET)$_4$ (Figure 1b) [1]. It is key to mutate positions near the active site to ensure their overall effect on catalytic activity [3]. Each residue position has the possibility of being filled with one of 20 canonical amino acids. In-silico computer simulations were first performed to screen for the large amount of possible *Is*PETase mutations.

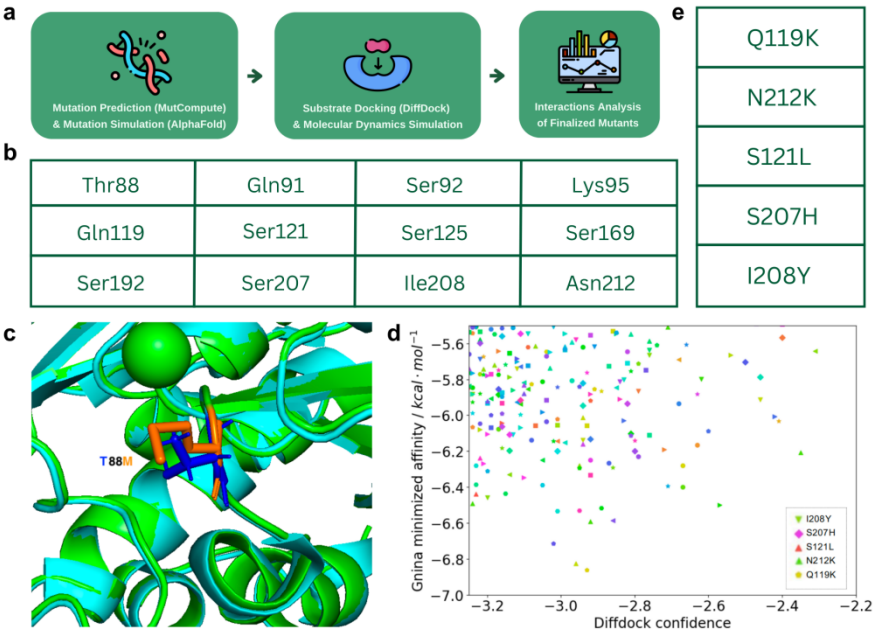

Figure 2. In-Silico Workflow was First Used to Screen for PETase Mutants. a. Simple flowchart of computer simulation softwares based on machine learning to select optimal PETase mutants. Initial optimal mutations are identified with MutCompute and screened further with docking simulations. The

DOI: 10.54254/2753-8818/67/2024.18882

stability of PETase is analyzed with molecular dynamics and new interactions are observed with protein visualization software. b. 12 selected mutants from MutCompute that can be predicted for gain-of-function mutations. Each position's original amino acid residue is labelled with its 3-letter code and position. c. Example AlphaFold mutation from the original threonine to methionine in the 88th position of the amino acid sequence. The original enzyme and the mutant are aligned to each other with the original threonine portrayed in blue while methionine is orange. d. Graph of DiffDock confidence to Gnina minimized affinity (kcal•mol^{-1}) for DiffDock result of each AlphaFold-simulated PETase mutant. The 5 predicted models of the same PETase mutant were assigned to the same series with a specific color and shape. Each individual point is plotted on the graph, whose window has been adjusted to show the most optimal mutants (high affinity and confidence). DiffDock confidence ranges from -3.25 to -2.2 while Gnina minimized affinity ranges from -7,0 to -5.5. e. 5 chosen PETase mutants for laboratory characterization. The first letter is the one-letter code of the original amino acid and the last letter is that of the amino acid it is mutated into. Residue position is described via the number in the middle.

3.2. Predicting Mutant PETase Viability

The entire in-silico computer simulation process illustrated in Figure 2a consists of the use of numerous computer software, many based in machine learning, to select mutants for later experimentation in the laboratory. This pathway can be split into two parts: predicting the improved function of various mutants to initially select mutants and analyzing mutants for finalization before lab work. Notably, the second phase is mainly for the verification of mutants and evaluation for what reasons they have been predicted to bring gain-of-function mutations. The mutants are not necessarily adjusted based on those results. The first phase consists of mutation prediction, mutation simulation, and substrate docking. Mutation prediction was completed using the 3D convolutional neural network tool MutCompute. Upon inputting the PDB ID 6EQE corresponding to the *Is*PETase protein, the software shortly returned the certainty that a position can provide a gain-of-function mutation to the entire protein. We chose 12 positions that have the lowest certainty for its current amino acid residue being the most optimal and are near the active site, to ensure its role in mainly catalytic activity (Figure 2b). Subsequently, we mutated each of the 12 positions once with each of the other 19 canonical amino acids, producing 228 total new mutants.

To better evaluate the effects of these mutants, Alphafold 2, a deep learning-based software, predicted the protein folding of the proteins into 3D structures. The code was run in Google Colaboratory and was optimized for bulk simulation that requires little human intervention during the long simulation time (Figure 2c shows the example of a mutation that evidently changed the protein's structure in this residue's case, especially).

As PET chains are extremely long in practice, various papers utilize different comparable models for its in-silico evaluation, like repeated oligomers of the components of PET. Longer oligomers have a higher resemblance to the behavior of the actual chain, but cost will increase as well. Here, this paper employs a PET quadromer (refer to monomer structure in Figure 1a) [3, 4]. Avogadro was used to create, equilibrate, and prepare the quadromer [9] with its SMILES notation O(C(=O)c1ccc(cc1)C(=O)OCCO)CCOC(=O)c1ccc(cc1)C(=O)OCCOC(=O)c1ccc(cc1)C(=O)OCCOC(=O)c1ccc(cc1)C(=O)O for downstream analysis.

With both the 3d structures of the ligand and receptor visualized, substrate-docking simulation allows us to determine whether the two components have an increase affinity to aid the catabolism of the PET ligand. DiffDock, a diffusion generative model, facilitated the simulation while another machine learning tool in Gnina further optimized the predicted binding affinity by applying its model based on physical principles. The reason this model was used is that diffusion generative model has offered more precision than the traditional and deep learning methods for ligand docking. AlphaFold returned 5 predicted models per folded protein, and each was run through DiffDock. As the software returned 40 predicted results per entry, each mutant ended up with 200 versions of predicted binding affinity and DiffDock confidence returned. To compile this large amount of data, Google Colaboratory and specifically the Matplotlib digital library was applied. All of the points for the same mutant were organized into the same series and a graph (Figure 2d) was plotted to allow the better identification of

DOI: 10.54254/2753-8818/67/2024.18882

optimal mutants. Its window only displayed mutants with a DiffDock confidence higher than -3.25 and a Gnina minimized affinity lower than -5.5 (indeed, a lower affinity value means higher affinity). Mutants were chosen to have especially low affinity values (like Q119K, which had the lowest affinity value out of all mutants in the acceptable docking confidence range). Consistency was also key: mutants like N212K, which appeared three times with affinities below -6.1 across different DiffDock confidence levels, were flagged as potentially reliable. Using this method, 5 mutants were ultimately chosen: Q119K, N212K, S121L, S207H, I208Y (Figure 2e). Although many residues at the same position produced excellent results, the chosen mutants were ensured to have different positions with the goal of finding the best mutation for a single position.

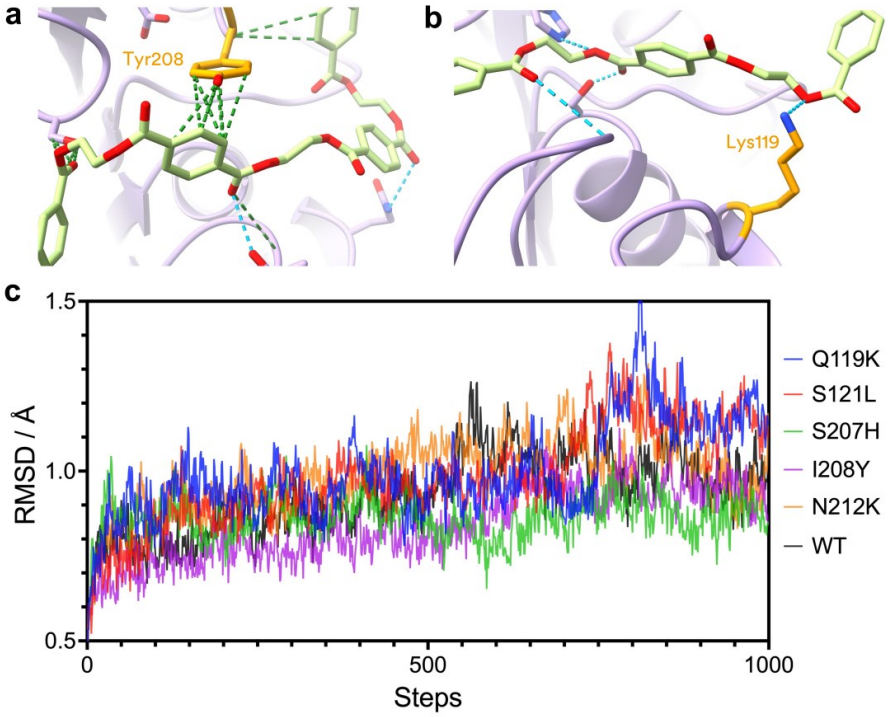

Figure 3. Chosen PETase Mutants were Analyzed for New Interactions and Stability. a. Depiction produced by ChimeraX of interactions between tyrosine at the 208th position with the PETase quadromer. The orange amino acid represents the tyrosine residue, and the green ligand is the PETase quadromer. The dark green lines in between display a π--π stacking interaction formed. b. Depiction produced by ChimeraX of interactions between lysine at the 119th position with the PET quadromer. The color labels are the same with lysine being orange and the ligand shown in green. The blue interaction in between entails a hydrogen bond formed. c. Line graph of root mean square deviation (Å) to steps during molecular dynamics simulation of 6 series, each representing a unique mutant protein. The 1000 steps make up the total simulation time during which RMSD is an indicator of the spatial difference between the original positions of alpha carbons to its changed location as it interacts with water molecules within a box system. It acts as an evaluator of protein stability with low RMSD representing high stability and vice versa.

3.3. Analyzing and Finalizing Chosen Mutants

Before sending the 5 proteins to experimental characterization, we applied several in-silico analysis techniques that are not machine learning-based to evaluate these enzymes. While machine learning is great at predictions, it can be insightful to evaluate whether there are any tangible improvements to the mutants based on established methods. Using the molecular visualization software ChimeraX and its structural analysis tools to identify bonds, two PETase mutants were found to produce new bonds with

DOI: 10.54254/2753-8818/67/2024.18882

the PET quadromer when docked as the location predicted by DiffDock. For the mutation I208Y at position 208 where the residue is mutated from isoleucine to tyrosine, Figure 3a depicts a green interaction between the PETase mutant and the PET quadromer. Its complex array of lines illustrating interactions and the fact that the interaction is occurring between two aromatic components suggest that this is a new π-π stacking interaction. Isoleucine is not aromatic; therefore, it is only because of the mutation to tyrosine that this new interaction is formed at position 208. At position 119, the chosen mutant Q119K mutates glutamine into lysine. The blue line in Figure 3b demonstrates the formation of a hydrogen bond. While glutamine is uncharged, the charged property of lysine led to the newly formed interaction. These newly formed interactions from PETase mutants add rationale to why they were chosen as interactions can increase stability and binding affinity [3, 4], in turn enhancing catalytic activity. For the other mutants, it is hypothesized that they experience the distal effect [4]. Even though no new interactions are apparent, the mutation likely changed their overall structure to improve affinity and catabolism.

In addition to interactions analysis, molecular dynamics simulation with the OpenMM toolkit applied in Google Colaboratory was also conducted with the mutants to analyze their stability in a dynamic system. The root mean square deviation was of particular focus, as it represents the overall change of the alpha carbons within of the protein and ligand inputted into the system. High RMSD represents low stability while low RMSD represents high stability. Since RMSD is relative, mutants could be compared to the wild-type for evaluation. From general trends across time, it is observed that the mutants S207H and I208Y performed the best having relatively stable RMSD that was lower than the wild-type, promising potential in producing positive experimental results [23]. N212K and S121L had a similar RMSD to the WT. The largest spike in RMSD was observed in the mutant Q119K where its RMSD reached 1.5 Å at one point. However, Figure 3b corroborates the fact that the mutation causes the formation of new interactions between the enzyme and the ligand, increasing affinity. Therefore, the mutant Q119K was sent alongside the rest of the mutants for experimental characterization regardless. How the mutant's low stability but high binding affinity affected the mutant's overall performance was a topic of curiosity.

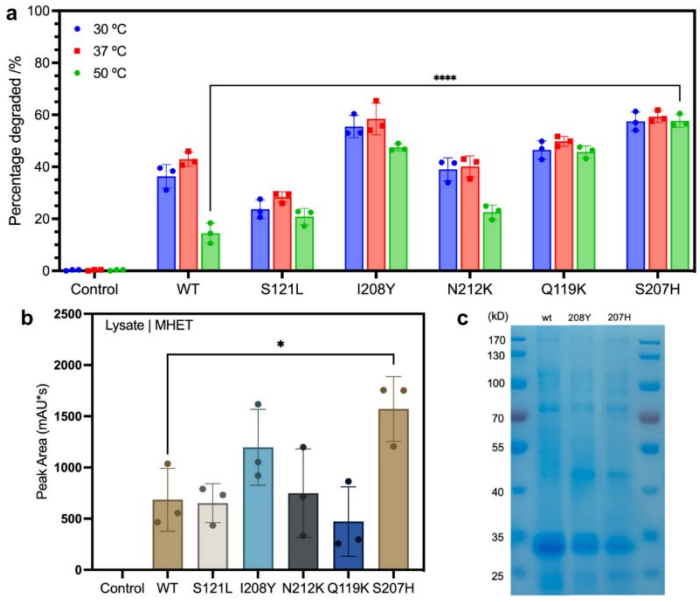

Figure 4. Chosen Mutants were Assessed as Cell Lysate with Optimal Mutants later being Chosen for Protein Expression. a. Bar graph of different PETase mutants to the percentage of PET degraded measured based on change in mass. A negative control was applied in having pure water incubated with the PET granules. 3 temperature conditions were tested for each mutant: 30° C, 37° C, 50° C. Each test case was completed in triplicates and standard deviation is plotted. Significance is plotted based on the

DOI: 10.54254/2753-8818/67/2024.18882

Student's t-test. b. Bar graph of peak area of MHET detected by HPLC assay that is produced from PET degradation for each mutant at 37° C. The negative control was once again a pure water solution. The temperature condition was kept at 37° C and each test case was done in triplicates to calculate standard deviation. Significance is plotted based on the Student's t-test. **c.** Picture of SDS-PAGE gel dyed with Coomassie Blue dye showing the purified amount of the wild-type *Is*PETase and the mutants I208Y and S207H. The very left and right column are protein ladders to show location of proteins of various masses. *Is*PETase has a molecular mass of ~30 kD.

3.4. Preliminary Testing of Catalytic Activity with Protein Cell Lysate

The 5 mutants selected from in-silico screening were then brought into the laboratory to test their viability through experiments. Aiming for a high-efficiency method of mutant screening, mutants were first sent to testing with their cell lysate [9]. Two main ways were utilized to achieve characterization of the cell lysates of mutants: mass analysis and HPLC assay. For each test case of a PETase protein, 3 trials were conducted for each of the 3 temperature conditions (30° C, 37° C, 50° C). After the incubation of 2 ml cell lysate with 50 mg PET granules for 5 days, the portion of PET granules that is yet to be degraded is taken for mass analysis by measuring weight and obtaining the percentage degraded. Depicted in Figure 4a, S121L and N212K showed a similar degradation efficiency to the wild-type enzyme. For the mutant Q119K, its degradation performance at 30° C and 37° C increased slightly, and the percentage of PET degraded at 50° C experienced a 3-fold increase, demonstrating the thermostability improvements made to the enzyme from the mutation. Mutants S207H and I208Y had the highest improvements in catalytic activity. While the wild-type enzyme degraded 36.3%, 42.9%, and 14.5% of the PET granules, mutant S207H caused the degradation of 57.5%, 59.3, and 57.7% of the PET under the incubation temperatures of 30° C, 37° C, and 50° C, respectively. An approximate 1.5-fold increase is seen in degradation effectiveness at 30° C and 37° C as well as a 4-fold increase in activity at 50° C with p-value calculated of less than 0.0001 to indicate the result's statistical significance.

The second approach used to evaluate the viability of the cell lysate of mutants was through HPLC assay. A 20 μL sample was taken from the reaction mixture of each trial where PETase was incubated with the PET granules at 37° C, which was used as an input to the HPLC machine. According to the peak area at the MHET retention time which suggests the abundance of the degraded products (shown in Figure 4b), the mutants S121L, N212K, and Q119K had similar or even worse activity compared to that of the wild-type protein. I208Y exhibited an impressive increase in degradation activity as its peak area detected for the MHET product was 1196.8 mAU*s compared to the wild-type PETase which had a peak area of 685.2 mAU*s. Meanwhile mutant S207H showed the largest improvement in activity once again with a peak area of 1571.5 mAU*s. This is more than a 2-fold increase in activity. Its significance is verified once again by the Student's t-test. However, the p-value is not as low as that of the mass analysis, largely due to the inconsistency of data within the HPLC assay. Overall, it is evident that the two mutants that displayed the best performance in degrading PET throughout the cell lysate experiments were S207H and I208Y.

3.5. Characterization of Performance using Purified Protein of Mutant

Referencing the most optimal mutants in preliminary cell lysate testing, mutants S207H and I208Y subsequently underwent histidine-tag protein purification alongside the wild-type for further verification of results. In Figure 4c, the purified proteins underwent SDS-PAGE for protein characterization. At 30 kD [24], a clear band is seen for each protein, highlighting the successful purification of all three proteins.

With the purification of the proteins, it was much more convenient to balance our protein abundance in each solution. Using BCA assay, protein concentration was determined and the concentration for each solution was balanced with dilution. Subsequently, 1 mL of 1 mg/mL purified protein was added to each of the reaction mixtures with PET for incubation for 5 days. Both mass analysis and HPLC assay was conducted with the purified protein as well. For the mass analysis, incubation conditions were the following two: 37° C and 50° C. 3 trials were completed for each mutant under each temperature. Here, both mutants displayed excellent performance relative to the wild-type *Is*PETase, especially the mutant

DOI: 10.54254/2753-8818/67/2024.18882

S207H. At 37° C, there was a 1.3-fold increase in degradation activity when comparing the WT enzyme to the mutant S207H. Mutant I208Y followed in performance right after the other mutant. A large difference between the performance between the two mutants is observed with the enzymes' incubation with PET at 50° C. The wild-type enzyme here degraded 13.3% of the PET; mutant I208Y degraded 39.6% PET; and mutant S207H degraded 62.2% of the PET granules (Figure 5a). While both mutants had increased degradation of PET, S207H stood out to have an over 4.5-fold increase in degradation and an impressive improvement in thermostability, as the incubation occurred at 50° C. The statistical significance of the result is supported by the p-value < 0.0001. Meanwhile, I208Y still had an impressive near 3-fold activity increase.

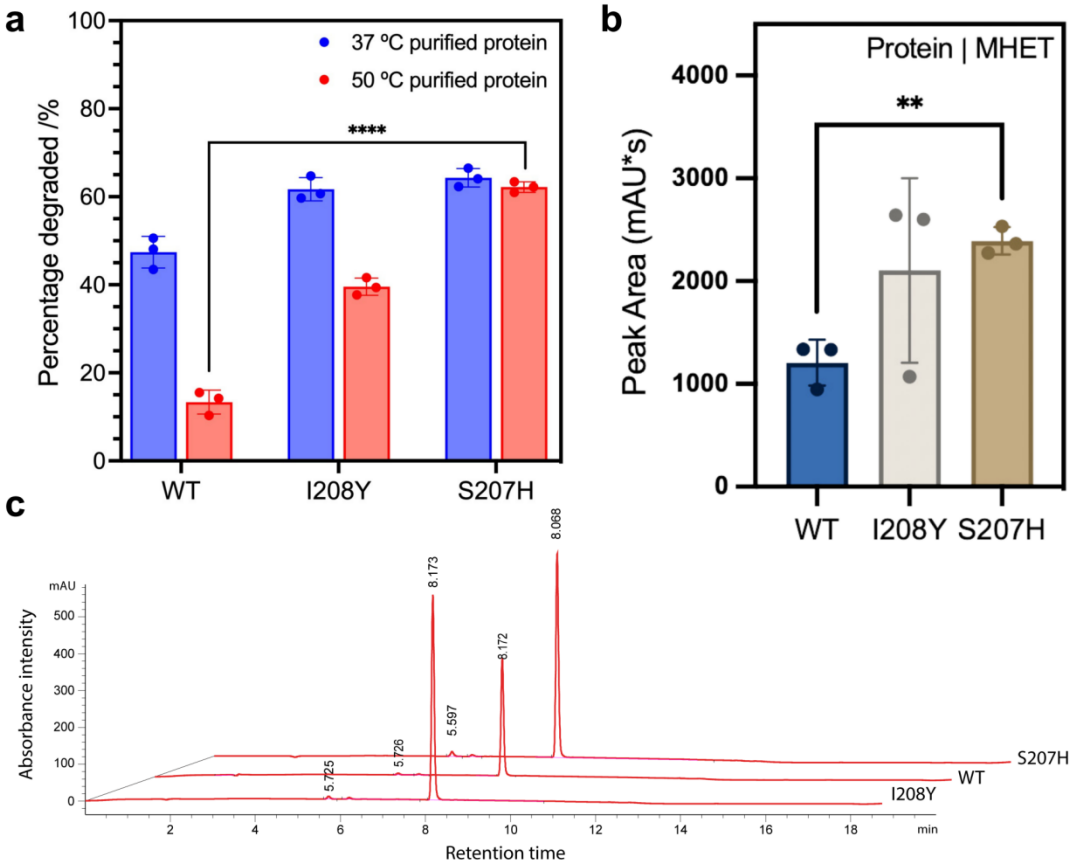

Figure 5. After protein purification, select mutants and wild-type were compared once again for performance. a. Bar graph of WT protein and mutants I208Y and S207H for the percentage of PET degraded observed via mass analysis. The purified proteins underwent two incubation conditions with PET granules at the temperature of 37 and 50° C. Each test case was done in triplicates and the standard deviation was plotted. Statistical significance is displayed on the graph and calculated via the Student's t-test. **b.** Bar graph of MHET produced from PET degradation detected from HPLC assay of WT and 2 mutants at 37° C. Tests are done in triplicates with standard deviation and statistical significance plotted (calculated with Student's t-test). **c.** Chromatogram from HPLC system plotting the absorbance intensity to retention time including the WT protein and mutants S207H and I207Y. The first peak labeled at ~5.725 min is that of TPA while the largest peak at ~8.173 min is that for MHET. Abundance of a compound is generally calculated via the area under the peak.

The solutions after incubation at the temperature condition of 37° C were once again utilized for HPLC assay. Mutant S207H had a peak area at the MHET retention time of 2391.4 mAU*s while the wild-type enzyme's peak area was 1206.6 mAU*s. I208Y trailed slightly behind S207H for its

DOI: 10.54254/2753-8818/67/2024.18882

degradation activity (Figure 5b). In addition, Figure 5c displays the chromatogram of the HPLC assay where the peak area has been calculated. This is another visual to demonstrate the mutant's overall improved performance (Figure 5c). From analyzing the results, a consistent 2-fold increase from the wild-type enzyme's activity to the mutants S207H and I207Y was seen, which is comparable to the HPLC assay with the cell lysate.

4. Conclusion

Protein engineering has long been a focus of the scientific community for optimizing protein effectiveness as well as preparing them for further use. While other methods such as directed evolution and rational engineering are feasible, there lacks a single solution to achieve both the screening of an abundance of mutants and the rational analysis of them without high cost. Here, we have developed a new machine learning-based solution to the protein engineering problem applicable to various proteins, specifically enzymes, in fields extending to sustainability, pharmaceuticals, and more [9]. This study demonstrates the use of this overall process that aims to drive up the efficiency of optimizing proteins, which is showcased via the engineering of *Is*PETase for its catalytic activity in degrading PET. With the machine learning-based in-silico computer simulations, we showed the screening of 228 different mutants to have been predicted for increased catabolic activity, which is narrowed down to only 5 mutants through a variety of rational docking and interaction simulations, like ligand docking simulation and molecular dynamics simulation. Experimental analysis using cell lysates and purified proteins confirmed that mutants S207H and I208Y consistently produced the most significant improvements in PET degradation. The most significant enhancement of PETase observed was indeed a 4.5-fold increase in degraded PET for mutant S207H as the purified protein degraded the prepared PET granules at 50° C. Moreover, it is interesting that thermostability improvements arise alongside the catalytic activity improvements, which is possible due to the algorithm of MutCompute to consider both factors in their predictions. While the fold increase may be higher for the mass analysis overall compared to HPLC assay, it is explainable due to the other compounds inside the equation like TPA (observed in small quantities in Figure 5c at an earlier retention time compared to that of MHET) and bis(2-Hydroxyethyl) terephthalate that is not accounted for.

Overall, the optimization of PETase holds a bright future. As the effectiveness of the biological degradation method in using PETase increases, there will naturally come opportunities for its integration for commercial use on a larger scale [9]. Moreover, recent discussions include the use of degraded PET products to achieve the production of virgin PET [4], achieving a cycle of recycling [25]. In the future, our machine learning-based protein engineering pathway can be optimized in many ways. It would be interesting to evaluate the effects of stacking mutations at different positions together to create one holistic mutant. One thing to evaluate is the mutations' total effect on catalytic activity and thermostability. Fortunately, many previous PET engineering papers have indeed incorporated many mutations to work together to devise their final mutant [9, 10, 25]. In regard to the machine learning portion of the process, much work is needed to improve the ligand docking analysis. There owes to be a balance between having repeated mutants that fit the criteria that we are looking for as well as maximum improvements in binding affinity. The development of a machine learning model or a numeric system to determine chosen mutants can greatly increase fairness and potentially rates of success. When choosing mutants from MutCompute, one also can only choose to evaluate a few specific amino acids at a position rather than mutating the position with all of the other canonical amino acids, which can leave space for the analysis of more plausible mutations. By doing so, not only will the protein engineering platform be of impact to the degradation of plastic waste, but it will also contribute to a larger community of proteins that can be engineered to solve many more problems in the world.

References

[1] Joo, S., Cho, I. J., Seo, H., Son, H. F., Sagong, H.-Y., Shin, T. J., Choi, S. Y., Lee, S. Y., & Kim, K.-J. (2018). Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation. Nature Communications, 9(1). https://doi.org/10.1038/s41467-018-02881-1

[2] Yang, H., Chen, G., & Wang, J. (2021). Microplastics in the marine environment: Sources, Fates, impacts and microbial degradation. Toxics, 9(2), 41. https://doi.org/10.3390/toxics9020041

[3] Han, X., Liu, W., Huang, J.-W., Ma, J., Zheng, Y., Ko, T.-P., Xu, L., Cheng, Y.-S., Chen, C.-C., & Guo, R.-T. (2017). Structural insight into catalytic mechanism of pet hydrolase. Nature Communications, 8(1). https://doi.org/10.1038/s41467-017-02255-z

[4] Austin, H. P., Allen, M. D., Donohoe, B. S., Rorrer, N. A., Kearns, F. L., Silveira, R. L., Pollard, B. C., Dominick, G., Duman, R., El Omari, K., Mykhaylyk, V., Wagner, A., Michener, W. E., Amore, A., Skaf, M. S., Crowley, M. F., Thorne, A. W., Johnson, C. W., Woodcock, H. L., ... Beckham, G. T. (2018). Characterization and engineering of a plastic-degrading aromatic polyesterase. Proceedings of the National Academy of Sciences, 115(19). https://doi.org/10.1073/ pnas.1718804115

[5] Jerves, C., Neves, R. P. P., Ramos, M. J., Silva, S. de, & Fernandes, P. A. (2021). Reaction mechanism of the PET degrading enzyme PETase studied with DFT/mm molecular dynamics simulations. ACS Catalysis, 11, 11626-11638. https://doi.org/10.1021/acscatal.1c03700.s002

[6] Singh, A., Rorrer, N. A., Nicholson, S. R., Erickson, E., DesVeaux, J. S., Avelino, A. F. T., Lamers, P., Bhatt, A., Zhang, Y., Avery, G., Tao, L., Pickford, A. R., Carpenter, A. C., McGeehan, J. E., & Beckham, G. T. (2021). Techno-economic, life-cycle, and socioeconomic impact analysis of enzymatic recycling of poly(ethylene terephthalate). Joule, 5(9), 2479–2503. https://doi.org/10.1016/j.joule.2021.06.015

[7] Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., Toyohara, K., Miyamoto, K., Kimura, Y., & Oda, K. (2016). A bacterium that degrades and assimilates poly(ethylene terephthalate). Science, 351(6278), 1196–1199. https://doi.org/10.1126/ science.aad6359

[8] Xu, Y., Cui, R., Han, Y., Jiang, J., Hu, D., Zhao, L., & Xi, Z. (2024). Efficient alcoholysis of poly(ethylene terephthalate) by using supercritical carbon dioxide as a green solvent. Polymers, 16(11), 1564. https://doi.org/10.3390/polym16111564

[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., Levy, C., Shaver, M. P., & Green, A. P. (2022). Directed evolution of an efficient and thermostable PET depolymerase. Nature Catalysis, 5(8), 673–681. https://doi.org/10.1038/s41929-022-00821-3

[10] 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., Lynd, N. A., Ellington, A. D., & Alper, H. S. (2022). Machine learning- aided engineering of hydrolases for PET depolymerization. Nature, 604(7907), 662–667. https:// doi.org/10.1038/s41586-022-04599-z

[11] Son, H. F., Cho, I. J., Joo, S., Seo, H., Sagong, H.-Y., Choi, S. Y., Lee, S. Y., & Kim, K.-J. (2019). Rational Protein Engineering of Thermo-Stable PETase from Ideonella sakaiensis for Highly Efficient PET Degradation. ACS Catalysis, 9(4), 3519–3526. https://doi.org/10.1021/acscatal.9b00568

[12] Shroff, R., Cole, A. W., Diaz, D. J., Morrow, B. R., Donnell, I., Annapareddy, A., Gollihar, J., Ellington, A. D., & Thyer, R. (2020). Discovery of novel gain-of-function mutations guided by structure-based Deep Learning. ACS Synthetic Biology, 9(11), 2927–2935. https://doi.org/10.1021/acssynbio.0c00345

[13] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

[14] Forli, S., Huey, R., Pique, M. E., Sanner, M. F., Goodsell, D. S., & Olson, A. J. (2016). Computational protein–ligand docking and virtual drug screening with the autodock suite. Nature Protocols, 11(5), 905–919. https://doi.org/10.1038/nprot.2016.051

[15] Corso, G., Stärk, H., Jing, B., Barzilay, R., & Jaakkola, T. (2022). DiffDock: Diffusion steps, twists, and turns for molecular docking. arXiv preprint arXiv:2210.01776.

[16] Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L.-P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., Wiewiora, R. P., Brooks, B. R., & Pande, V. S. (2017). OpenMM 7: Rapid development of high performance algorithms for Molecular Dynamics. PLOS Computational Biology, 13(7). https://doi.org/10.1371/journal.pcbi.1005659

[17] Meng, E. C., Goddard, T. D., Pettersen, E. F., Couch, G. S., Pearson, Z. J., Morris, J. H., & Ferrin, T. E. (2023). UCSF ChimeraX: Tools for structure building and analysis. Protein Science, 32(11). https://doi.org/10.1002/pro.4792

[18] Bisong, E. (2019). Google Colaboratory. Building Machine Learning and Deep Learning Models on Google Cloud Platform, 59–64. https://doi.org/10.1007/978-1-4842-4470-8_7

[19] Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., & Hutchison, G. R. (2012a). Avogadro: An advanced semantic chemical editor, visualization, and analysis Platform. Journal of Cheminformatics, 4(17). https://doi.org/10.1186/1758-2946-4-17

[20] McNutt, A. T., Francoeur, P., Aggarwal, R., Masuda, T., Meli, R., Ragoza, M., Sunseri, J., & Koes, D. R. (2021). GNINA 1.0: Molecular docking with deep learning. Journal of Cheminformatics, 13(1). https://doi.org/10.1186/s13321-021-00522-2

[21] Moruzzi, G. (2020). Plotting with matplotlib. Essential Python for the Physicist, 53–69. https://doi.org/10.1007/978-3-030-45027-4_3

[22] Kufareva, I., & Abagyan, R. (2012). Methods of protein structure comparison. Methods in molecular biology, 857, 231–257. https://doi.org/10.1007/978-1-61779-588-6_10

[23] Deng, J.-H., Luo, J., Mao, Y.-L., Lai, S., Gong, Y.-N., Zhong, D.-C., & Lu, T.-B. (2020). π-π stacking interactions: Non-negligible forces for stabilizing porous supramolecular frameworks. Science Advances, 6(2). https://doi.org/10.1126/sciadv.aax9976

[24] Jia, Y., Samak, N. A., Hao, X., Chen, Z., Wen, Q., & Xing, J. (2022). Hydrophobic cell surface display system of PETase as a sustainable biocatalyst for PET degradation. Frontiers in Microbiology, 13, 1005480. https://doi.org/10.3389/fmicb.2022.1005480

[25] Tournier, V., Topham, C. M., Gilles, A., David, B., Folgoas, C., Moya-Leclair, E., Kamionka, E., Desrousseaux, M.-L., Texier, H., Gavalda, S., Cot, M., Guémard, E., Dalibey, M., Nomme, J., Cioci, G., Barbe, S., Chateau, M., André, I., Duquesne, S., & Marty, A. (2020). An engineered PET depolymerase to break down and Recycle Plastic Bottles. Nature, 580(7802), 216–219. https:// doi.org/10.1038/s41586-020-2149-4