Machine Learning Applied to the Optimization of Plastic-Degrading Enzymes

Romeu Fernandes[1][4110−7A60−7D40]

University of Minho, Braga, Portugal

1 Introduction

1.1 The Global Plastic Crisis

Plastic pollution has become a major problem over the years and is degrading the quality of the surroundings worldwide. Plastic production has surged globally since the 1950s, reaching 380 million tons annually (1). Traditional plastics take centuries to decompose, worsening the problem (2).

A noteworthy fact is that, only 9% of plastic waste is recycled while the remaining percent is either burnt, or dumped into landfills where it continues to pollute the environment. This has led to nearly 8 million tons of plastic ending up in the oceans each year (3). As a result, this excess of plastic has created marine garbage patches around the world, the largest being near the Pacific Ocean, which is more than three times the size of France (4).

Often, marine animals get caught in plastic waste or swallow plastic pieces, resulting in injury, hunger, and eventually death (5). Microplastics - particles smaller than 5 mm - have spread through all levels of the food web and can be found in remote regions, from deep ocean trenches to Arctic ice (6). Recently, microplastics have been found in human blood, placenta, and lung tissue, raising substantial alarm for future health problems (7).

Moreover, traditional methods of producing plastic are very fossil fuel intensive, as almost 6% of global oil consumption is directed towards the production of plastic, which is expected to grow to 20% by 2050 (8). As such, the search for biological methods for the breakdown of plastic, especially by using enzymes, has become a key focus of research to address this multifaceted problem (9)(10).

1.2 Enzymatic Solutions for Plastic Degradation

The identification of plastic-degrading enzymes in a range of microorganisms has created an exciting new avenue in the biological management of plastic waste. Synthetic polymers, such as plastics, were previously believed to be extremely resistant to biological degradation but now can be cleaved into monomers and oligomers (11)(12).

The microbial sources of these enzymes are rather broad, including bacteria from plastic polluted environments like *Ideonella sakaiensis*, which produces PETase for degrading polyethylene terephthalate (PET), and numerous fungi

that produce cutinases that can hydrolyze several synthetic polymers (9)(13). The range of known plastic degrading enzymes includes polyethylene terephthalate hydrolases (PETases), cutinases, lipases, esterases, laccases, manganese peroxidases, and alkane hydroxylases, which all look for different types of plastics or certain chemical bonds within the plastic (14)(15).

These enzymes act through a variety of mechanisms: hydrolases such as PETase hydrolyze ester bonds in polyesters via hydrolysis, while oxidative enzymes (e.g., laccases and peroxidases) degrade polymer chains via free radical mechanisms (10). Catalytic processes usually occur by the initial generation of enzyme–substrate complexes on the plastic surface, which is then followed by nucleophilic attacks on susceptible bonds within the polymer backbone (16). Although, with reaction rates that are far too slow for cost-effective plastic waste disposal, their inherent catalytic efficiencies are sometimes insufficient for industrial-scale applications (17). Additionally, many natural plastic-degrading enzymes exhibit limited stability under the harsh conditions often encountered in waste processing facilities, such as elevated temperatures, extreme pH values, or the presence of inhibitory compounds (12).

In addition, the heterogeneity of plastic waste streams, which include, among others, various polymer types and polymer additives, as well as contaminants, adds more challenges to enzymatic degradation systems (18). These limitations emphasize the necessity of using enzyme engineering techniques to improve the functional properties of the natural plastic degrading enzymes for real applications in waste management and recycling systems.

1.3 Machine Learning Approaches in Enzyme Engineering

Not surprisingly, machine learning (ML) is a game changer in enzyme engineering, marking a shift towards data-driven optimization strategies away from traditional rational design and directed evolution approaches. Such a computational strategy enables tuning of enzyme traits while skipping a complete mechanistic interpretation of sophisticated structure-function correlations (19)(20).

Conventional approaches to enzyme engineering have largely relied on heuristic expertise or random mutagenesis and selection, which can be costly and highly limited.

Different ML architectures have been shown to be effective for enzyme engineering. Trained on existing experimental data, supervised learning approaches (e.g., random forests, support vector machines, neural networks) can predict the performance of an enzyme based upon sequence or structural features (21)(22). Deep learning approaches, particularly convolutional neural networks and graph neural networks, have shown to be particularly effective at modeling the intricate relationships between protein sequence, structure, and function (23)(24).

When it comes to plastic degrading enzymes, ML models have been used to identify optimal mutations to improve key properties such as catalytic efficiency, thermal stability and substrate specificity (25). Often these include sequence-derived features, such as properties of amino acids and evolutionary

data, structural datasets including distance maps, solvent accessibility and secondary structure elements and also molecular dynamics simulations (19)(26).

ML-based approaches are powerful for enzyme engineering as machine learning has these capabilities to find non-obvious patterns and correlations in experimental data that would not be apparent to humans. Such a capability dramatically speeds up the classical enzyme engineering process by predicting and virtually eliminating many of the variants that a given experiment would have required to be tested, enabling an efficient biocatalyst development that is faster and less resource-demanding (26)(27).

Moreover, their models can be further trained on more data, leading to an increasingly virtuous cycle of enzyme design, where each round of predictions and experimental validation allows for improved model accuracy for the next round of predictions.

1.4 Computational Solutions to Plastic Pollution

With production of 380 million tons annually and only 9% being recycled, the global plastic crisis calls for quick and radical solutions. Two interesting directions are intended to be investigated by this project: the use of machine learning techniques to enzyme engineering and biological degradation using plastic-degrading enzymes. Through hydrolysis or oxidative processes, naturally occurring enzymes including PETase from *Ideonella sakaiensis* and several fungal cutinases have shown the capacity to break resistant synthetic polymers. Still, these natural enzymes have major drawbacks, particularly with relation to their catalytic efficiencies for uses in industry, dependability under limited and hostile process conditions, and challenges presented by heterogeneous waste streams including various types of polymers and additives.

From conventional rational design and directed evolution approaches towards data-driven optimization strategies, the integration of machine learning marks a paradigm change. ML approaches hasten the creation of strong biocatalysts by predicting beneficial mutations and effectively navigating the large sequence space of possible enzyme variants. Models of the complex interactions between protein sequence, structure, and function have shown success using many ML architectures including supervised learning techniques and deep learning methods.

Combining biotechnology with computational techniques applies the inherent benefits of both areas to possibly overcome the constraints of conventional plastic waste management and generate a virtuous cycle of enzyme design in which every round of predictions and experimental validation increases the model's accuracy. This project targets multidisciplinary strategies to tackle tough environmental issues related with plastic waste management.

2 Project Objectives and Aims

2.1 Primary Objective

Develop and implement a machine learning framework for identifying, optimizing, and characterization of novel plastic-degrading enzymes with increased catalytic efficiency and substrate specificity.

2.2 Specific Aims

Database Development and Knowledge Integration

- Improve the Plastizyme database of known plastic-degrading enzymes by integrating data from UniProt, NCBI Protein, PDB, and AlphaFold.
- Standardize and pre-process enzyme sequence and structural data to ensure quality and consistency for downstream machine learning applications.

Predictive Model Development

- Design and train machine learning models to predict enzyme functionality based on sequence and structural features.
- Compare performance of advanced deep learning approaches with classical machine learning methods to identify optimal modeling strategies for plastic-degrading enzyme prediction.
- EnzyNet: Develop a machine learning approach to predict enzyme function based on 3D structural features using deep learning models.
- UniRep: Leverage sequence-based protein embeddings to optimize function and stability predictions for protein engineering applications.

Structure-Function Relationship Analysis

- Clarify the structure-function relationships of plastic-degrading enzymes via computational modeling and docking simulations.
- Use AlphaFold3 software to create precise structural representations of promising enzyme candidates.
- Implement automated HADDOCK docking workflows for evaluating enzyme-substrate interactions with different plastics polymers.

Integrated Computational Pipeline Development

- Develop a seamless computational pipeline that uses sequence analysis, structural prediction, molecular docking, and machine learning to find and verify high-potential enzyme variants.
- Incorporate docking parameters and binding affinity data into machine learning models to improve forecast accuracy of enzyme activity.

Performance Validation and Optimization

- Systematically evaluate the predictive performance of developed models using Orange Data Mining software and suitable validation metrics.
- Maximize accuracy in detecting enzymes with increased catalytic activity by optimizing model parameters.

Biotechnological Application Assessment

- Evaluate the potential of identified enzyme candidates for practical applications in industrial plastic recycling operations and environmental bioremediation.
- Prioritize enzyme variants based on predicted efficiency, stability, and feasibility for scaled production.

This project is meant to tackle the international plastic pollution crisis by leveraging computational approaches to accelerate the discovery and optimization of enzymes capable of degrading various plastic polymers. By combining machine learning with structural biology approaches, one will have a methodical system for identifying enzyme candidates for experimental confirmation and later biotechnological application.

3 Methodologies

To find and optimize new plastic-degrading enzymes, this project will use a multi-step computational approach that combines data mining, machine learning, structural modeling, and molecular docking. The methodologies are organized as follows:

3.1 Database Development and Knowledge Integration

Data Aggregation, Standardization and Pre-processing:

- Gather structural and sequence data about enzymes from reputable databases like AlphaFold, PDB, NCBI Protein, and UniProt. The Plastizyme database will be updated and expanded as a result.
- Create standardized pipelines to pre-process, clean, and annotate the data that has been gathered. Reliable downstream machine learning applications are made possible by this step, which guarantees consistency and quality.

3.2 Predictive Model Development

Feature Extraction:

- Key features can be extracted from structural and sequence data. This includes extracting 3D structural attributes and using UniRep to generate protein embeddings.

Model Training, Comparison and Optimization:

- Using EnzyNet, create and deploy deep neural network topologies that capture intricate non-linear correlations between structural characteristics and enzyme activity.
- Create baselines and determine the best modeling approaches by developing and contrasting classical algorithms.
- To increase the models' generalization and predictive accuracy, apply cross-validation and hyperparameter tuning strategies.

3.3 Structure-Function Relationship Analysis

Computational Modeling and Molecular Docking:

- Use AlphaFold3 as well as Modeller to create precise 3D models of potential enzymes.
- To clarify important structure-function relationships, incorporate automated docking workflows with HADDOCK to model enzyme-substrate interactions with different plastic polymers.

3.4 Integrated Computational Pipeline Development

Pipeline Architecture, Data Fusion and Workflow Automation:

- Create a seamless, comprehensive computational pipeline that combines molecular docking, sequence analysis, structural prediction, and machine learning model inference.
- To improve predictions of enzyme activity, integrate docking parameters and binding affinity data into the machine learning framework.
- Automate model execution and data processing to guarantee reproducibility and make large-scale enzyme variant screening easier.

3.5 Performance Validation and Optimization

Validation Metrics:

- Analyze each model's predictive performance methodically using the visualization tools and established metrics in Orange Data Mining software.
- Perform iterative cycles of model training, validation, and parameter optimization to increase the precision of identifying enzymes with improved catalytic efficiency.

3.6 Biotechnological Application Assessment

Candidate Prioritization and Feasibility Analysis:

- Sort enzyme variations according to their expected scalability, stability, and efficiency for environmental bioremediation and industrial plastic recycling.
- Analyze the practicality of the best candidates using in silico tests that replicate industrial process conditions, offering a guide for further experimental verification.

Bibliography

[1] Geyer, R., Jambeck, J. R., & Law, K. L. (2017). Production, use, and fate of all plastics ever made. Science Advances, 3(7), e1700782. https://doi.org/10.1126/sciadv.1700782

[2] Barnes, D. K. A., Galgani, F., Thompson, R. C., & Barlaz, M. (2009). Accumulation and fragmentation of plastic debris in global environments. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1526), 1985-1998. https://doi.org/10.1098/rstb.2008.0205

[3] Jambeck, J. R., Geyer, R., Wilcox, C., Siegler, T. R., Perryman, M., Andrady, A., Narayan, R., & Law, K. L. (2015). Plastic waste inputs from land into the ocean. Science, 347(6223), 768-771. https://doi.org/10.1126/science.1260352

[4] Lebreton, L., Slat, B., Ferrari, F., Sainte-Rose, B., Aitken, J., Marthouse, R., Hajbane, S., Cunsolo, S., Schwarz, A., Levivier, A., Noble, K., Debeljak, P., Maral, H., Schoeneich-Argent, R., Brambini, R., & Reisser, J. (2018). Evidence that the Great Pacific Garbage Patch is rapidly accumulating plastic. Scientific Reports, 8(1), 4666. https://doi.org/10.1038/s41598-018-22939-w

[5] Wilcox, C., Van Sebille, E., & Hardesty, B. D. (2015). Threat of plastic pollution to seabirds is global, pervasive, and increasing. Proceedings of the National Academy of Sciences, 112(38), 11899-11904. https://doi.org/10.1073/pnas.1502108112

[6] Bergmann, M., Mützel, S., Primpke, S., Tekman, M. B., Trachsel, J., & Gerdts, G. (2019). White and wonderful? Microplastics prevail in snow from the Alps to the Arctic. Science Advances, 5(8), eaax1157. https://doi.org/10.1126/sciadv.aax1157

[7] Leslie, H. A., van Velzen, M. J. M., Brandsma, S. H., Vethaak, A. D., Garcia Vallejo, J. J., & Lamoree, M. H. (2022). Discovery and quantification of plastic particle pollution in human blood. Environment International, 163, 107199. https://doi.org/10.1016/j.envint.2022.107199

[8] World Economic Forum. (2016). The new plastics economy: Rethinking the future of plastics. World Economic Forum. https://www.weforum.org/reports/the-new-plastics-economy-rethinking-the-future-of-plastics

[9] 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

[10] Carr, C. M., Clarke, D. J., & Dobson, A. D. W. (2020). Microbial polyethylene terephthalate hydrolases: current and future perspectives. Frontiers in Microbiology, 11, 571265. https://doi.org/10.3389/fmicb.2020.571265

[11] Danso, D., Schmeisser, C., & Streit, W. R. (2019). Plastic biodegradation: Biotechnology and microbiology. Applied and Environmental Microbiology, 85(16), e01095-19. https://doi.org/10.1128/AEM.01095-19

[12] Wei, R., Song, C., Gräsing, D., Schneider, T., Bielytskyi, P., Böttcher, D., Matysik, J., Bornscheuer, U. T., & Zimmermann, W. (2019). Conformational fitting of a flexible oligomeric substrate does not explain the enzymatic PET degradation. Nature Communications, 10(1), 5581. https://doi.org/10.1038/s41467-019-13492-9

[13] Shirke, A. N., White, C., Englaender, J. A., Zwarycz, A., Butterfoss, G. L., Linhardt, R. J., & Gross, R. A. (2018). Stabilizing leaf and branch compost cutinase (LCC) with glycosylation: mechanism and effect on PET hydrolysis. Biochemistry, 57(7), 1190-1200. https://doi.org/10.1021/acs.biochem.7b01189

[14] Taniguchi, I., Yoshida, S., Hiraga, K., Miyamoto, K., Kimura, Y., & Oda, K. (2019). Biodegradation of PET: Current status and application aspects. ACS Catalysis, 9(5), 4089-4105. https://doi.org/10.1021/acscatal.8b05171

[15] Palm, G. J., Reisky, L., Böttcher, D., Müller, H., Michels, E. A. P., Walczak, M. C., Berndt, L., Weiss, M. S., Bornscheuer, U. T., & Weber, G. (2019). Structure of the plastic-degrading Ideonella sakaiensis MHETase bound to a substrate. Nature Communications, 10(1), 1717. https://doi.org/10.1038/s41467-019-09326-3

[16] 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), 382. https://doi.org/10.1038/s41467-018-02881-1

[17] 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

[18] Salvador, M., Abdulmutalib, U., Gonzalez, J., Kim, J., Smith, A. A., Faulon, J. L., Wei, R., Zimmermann, W., & Jimenez, J. I. (2019). Microbial genes for a circular and sustainable Bio-PET economy. Genes, 10(5), 373. https://doi.org/10.3390/genes10050373

[19] Yang, K. K., Wu, Z., & Arnold, F. H. (2019). Machine-learning-guided directed evolution for protein engineering. Nature Methods, 16(8), 687-694. https://doi.org/10.1038/s41592-019-0496-6

[20] Xu, Y., Verma, D., Sheridan, R. P., Liaw, A., Ma, J., Marshall, N. M., McIntosh, J., Sherer, E. C., Svetnik, V., & Voigt, J. H. (2021). Deep dive into machine learning models for protein engineering. Journal of Chemical Information and Modeling, 60(6), 2773-2790. https://doi.org/10.1021/acs.jcim.0c00073

[21] Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J., & Arnold, F. H. (2021). Machine learning-assisted directed protein evolution with combina-

torial libraries. Proceedings of the National Academy of Sciences, 118(10), e2012588118. https://doi.org/10.1073/pnas.2012588118

[22] Fox, R. J., Davis, S. C., Mundorff, E. C., Newman, L. M., Gavrilovic, V., Ma, S. K., Chung, L. M., Ching, C., Tam, S., Muley, S., Grate, J., Gruber, J., Whitman, J. C., Sheldon, R. A., & Huisman, G. W. (2007). Improving catalytic function by ProSAR-driven enzyme evolution. Nature Biotechnology, 25(3), 338-344. https://doi.org/10.1038/nbt1286

[23] Alley, E. C., Gimpel, J. A., Davies, J. H., & Colwell, L. J. (2019). Unified rational protein engineering with sequence-based deep representation learning. Nature Methods, 16(12), 1315-1322. https://doi.org/10.1038/s41592-019-0598-1

[24] Amidi, A., Amidi, S., Vlachakis, D., Megalooikonomou, V., Paragios, N., & Zacharaki, E. I. (2022). EnzyNet: enzyme classification using 3D convolutional neural networks on spatial representation. PeerJ, 10, e13380. https://doi.org/10.7717/peerj.4750

[25] Chen, Z., Wang, Z., Ren, J., Li, G., & Jiang, F. (2021). Machine learning-guided directed evolution for protein engineering. ACS Synthetic Biology, 10(12), 3141-3159. https://doi.org/10.1021/acssynbio.1c00297

[26] Mazurenko, S., Prokop, Z., & Damborsky, J. (2020). Machine learning in enzyme engineering. ACS Catalysis, 10(2), 1210-1223. https://doi.org/10.1021/acscatal.9b04321

[27] Kurinov, A., Gaetani, M., Mondal, A., & Greener, J. G. (2023). Advances and limitations of protein engineering with machine learning. Nature Communications, 14(1), 1-12. https://doi.org/10.1038/s41467-023-39796-5