NEXT-GEN ENZYME ENGINEERING – A WET LAB DATA-DRIVEN APPROACH TO IDENTIFY AND RECOMBINE KEY POINT MUTATIONS WITH ENZYMAP AI AND ENZYREC AI FOR SUPERIOR ENZYME PERFORMANCE

David SCHOENAUER, Aminoverse B.V.
schoenauer@aminoverse.com
Dr. Eduardo OLIVEIRA, Aminoverse B.V.
Kerstin CARDINIER, Aminoverse B.V.

Key Words: enzyme engineering, machine learning, alcohol dehydrogenase, ai

Directed evolution of enzymes has evolved towards increasing reliance on in silico designs and lessened dependence on expensive experimental screens. In the last few years we are witnessing the emergence of in silico Machine learning (ML) strategies with the promise to propel the field of enzyme evolution further, eventually reaching the goal of function-guided de novo enzyme design[1]. Aminoverse seeks to fulfill that promise with its proprietary EnzyMAP AI and EnzyREC AI.

EnzyMAP AI is a modern deep learning architecture that uses a small wet-lab data input of 3 to 5 representative amino acids, combined with evolutionary and structural information, to predict the fitness of all 20 amino acid mutations in all positions of any given enzyme. For training we collected and generated datasets with >150,000 mutants spanning diverse proteins.

With a wet lab data input of only 20%, EnzyMAP AI predicts the best performing point mutations (hotspots) across the entire enzyme sequence with >90% accuracy visualized by a fitness heatmap (Fig.1). This approach is especially useful to gather holistic knowledge about the effect of mutations towards a variety of enzyme properties while reducing screening efforts and costs by up to 85%.

Guided by the knowledge base generated via EnzyMAP AI, EnzyREC AI employs the latest advancements in large language models[2] and protein generative models conditioned on backbone coordinates[3,4] to cherry-pick and efficiently recombine key point mutations in final variants exhibiting superior enzyme performance.

We exemplify the synergies of EnzyMAP AI and EnzyREC AI with an internal case study revolving around an Alcohol Dehydrogenase engineered towards enhanced thermostability and activity. The performance of Aminoverse's AI platform is benchmarked against traditional approaches such as random mutagenesis and *in silico* design[5] and more recent publicly reported machine learning algorithms .

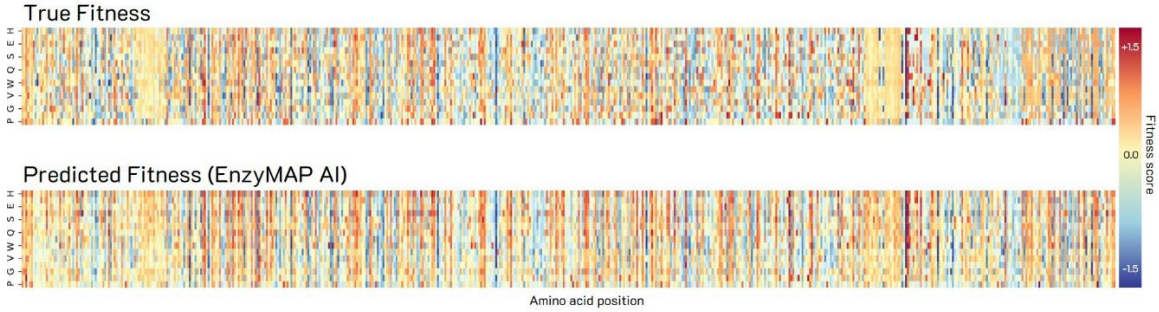

Figure 1: Fitness landscape prediction of EnzyMAP AI (bottom) on unseen experimental data set (top). Red color marks fitness improving mutations (hotspots), blue color marks fitness decreasing mutations (coldspots), yellow color indicates mutations not affecting fitness.

[1] Wittmann *et al.* 2021, "Advances in Machine Learning for Directed Evolution." https://doi.org/10.1016/j.sbi.2021.01.008.
[2] ESM – Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C. L.; Ma, J.; Fergus, R. Biological Structure and Function Emerge from Scaling Unsupervised Learning to 250 Million Protein Sequences. Proc. Natl. Acad. Sci. U.S.A. 2021, 118 (15), e2016239118. https://doi.org/10.1073/pnas.2016239118
[3] ProteinMPNN - Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Courbet, A.; de Haas, R. J.; Bethel, N.; Leung, P. J. Y.; Huddy, T. F.; Pellock, S.; Tischer, D.; Chan, F.; Koepnick, B.; Nguyen, H.; Kang, A.; Sankaran, B.; Bera, A. K.; King, N. P.; Baker, D. Robust Deep Learning–Based Protein Sequence Design Using ProteinMPNN. Science 2022, 378 (6615), 49–56. https://doi.org/10.1126/science.add2187
[4] MSATransformer - Rao, R.; Meier, J.; Sercu, T.; Ovchinnikov, S.; Rives, A. Transformer Protein Language Models Are Unsupervised Structure Learners; preprint; Synthetic Biology, 2020. https://doi.org/10.1101/2020.12.15.422761.
[5] FRESCO - Wijma, H. J.; Floor, R. J.; Jekel, P. A.; Baker, D.; Marrink, S. J.; Janssen, D. B. Computationally Designed Libraries for Rapid Enzyme Stabilization. Protein Engineering Design and Selection 2014, 27 (2), 49–58. https://doi.org/10.1093/protein/gzt061