3x Enzyme Productivity Improvement with No Prior Data

BioLM designs increase stability and productivity at +10°C compared to parent

SUMMARY

The founders of an early-stage biotech developing cell-free enzyme cascades needed to quickly improve the thermostability and catalytic activity of a lipase in Rounds 1-3. They were looking for an AI partner to provide candidate generation and lead selection based on their functional parent enzyme. Combining human expertise with proprietary machine learning pipelines, BioLM generated 200 sequences with distance N=1-10 from the parent to help the client achieve 3x activity improvement at +10°C in the first round of wet-lab testing, with no prior experimental data. All of the sequences were fully expressed, minimizing wasted cost on synthesis and testing. The client achieved their goal of improving the parent enzyme for bioreactor conditions, testing new and diverse sequences that were undiscovered by their rational designs.

HIGHLIGHTS

- Parent 0.4x productive at +10°C
- 100 optimized variants designed
- 100 *de-novo* leads generated
- 55% of variants have greater activity and stability at +10°C
- Ensemble of protein LLMs used

THE CUSTOMER

The Company is a US-based biotechnology startup with seed funding and less than 20 employees that utilizes its proprietary technology for ultra-stable biocatalyst production. These biocatalysts are developed for the Company's clients for carbon capture, waste breakdown, and more. When the Company engaged with BioLM, they had minimal wet-lab data to provide relating to activity and Tm, and were unsure if they could benefit from any AI because of their unique platform. They had read papers on AI-guided optimization of enzymes, but their expertise is as bench scientists. The Company was also conscientious of price and time, wanting results in a few weeks at less than the cost of a screen with a variant library.

CHALLENGES & OBJECTIVES

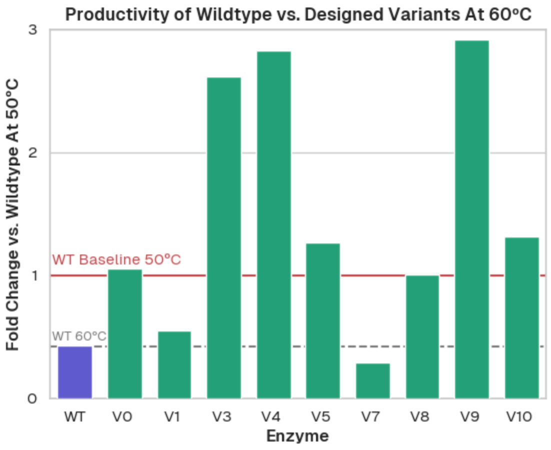
The Company needed to demonstrate incremental improvements in enzymatic stability, activity, and thermochemical tolerance to their client. They also wanted to use AI to identify novel sequences with the potential to achieve their goals and potentially reduce costs by minimizing rounds of wet lab testing. Given that they were in early stages of identifying a parental candidate, they only had wet lab data on a handful of enzymes, and were able to collect quantitative data to validate future recommended sequences.

THE IMPLEMENTATION

Changes in activity and Tm generally correlate with an increasing number of mutations to the parent. However, the Company was concerned about the risk of diverging too far from the parent sequences. BioLM suggested a hedged approach: in order to achieve the Company's dual objectives, 100 sequences with 1x-3x mutations from their parent were recommended as having greater likelihood of incremental improvements; and an additional 100 sequences with up to 10x mutations were generated to screen for large improvements in Tm and activity.

In the first phase, BioLM used an ensemble of 8 protein models to perform lead optimization with N=1-3 mutations from the parent enzyme. In the second phase, BioLM finetuned (or trained) a generative model for lead generation of stable lipases and screened millions of candidates *in-silico* up to N=10 mutations from the parent. To address the scarcity of data on the vast sequence search space, the training data was augmented with the topmost likely sequences from *in-silico* deep mutational scanning (DMS) of N=1-3. In total, 200 sequences were delivered to the client with an expected 30-40% of variants to improve upon the parent in multiple areas. The models used to screen the *in-silico* candidates involve ddG (ΔΔG), Tm, likelihood of function, expressibility, functional prediction, structural prediction, and more. The BioLM team included ML engineers, a bioinformatics scientist, and a data scientist.

THE RESULT



The client was happy to receive sequences with the active site and key regions conserved. Additionally, they were impressed to see that BioLM suggested mutations to positions they had previously considered in their rational designs. The client ordered and synthesized the top 10 sequences with N=1-3 mutations, then tested both at 50°C and 60°C in their in-house lab. While the parent enzyme performed worse at 60°C than 50°C, nearly all the sequences BioLM delivered performed better than the parent at 60°C. More than 80% were functional, and several of these 10 delivered candidates performed better at 60°C than 50°C, including one that demonstrated a *3x gain in productivity at this greater temperature*. The Company retained their IP and ownership of these sequences.