

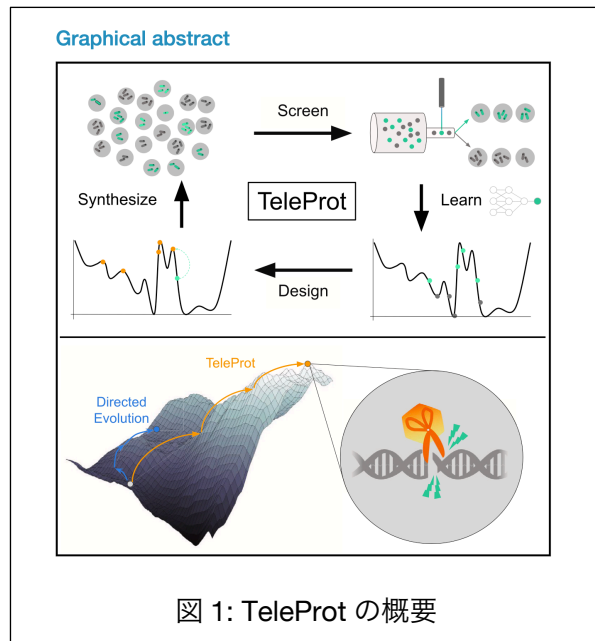
Journal Club (2025 年 4 月 21 日)

発表論文: Engineering highly active nuclease enzymes with machine learning and high-throughput screening

Thomas et al., Cell Systems 2025

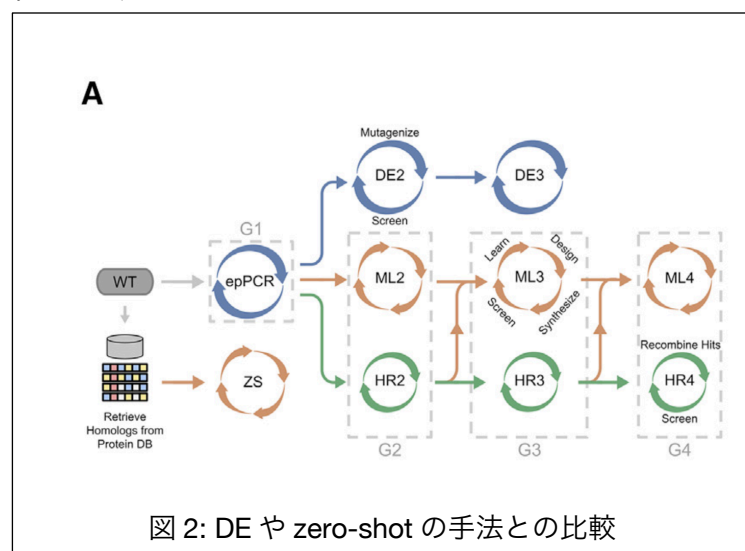
研究目的および概要

本研究は、創傷部位など生理的 pH7 で作用可能なバイオフィルム分解ヌクレアーゼ NucB (至適 pH9、pH7 では活性が 80%低下する) を高活性化することを目標とする。これを達成するために、筆者らは実験高速実験系と進化情報を組み込んだモデルを統合した TeleProt フレームワークを構築し、機械学習でライブラリ設計→超高スループットで評価→モデル再学習...というループを回して最適化した (図 1)。その結果、pH7 において WT 比で 19 倍の活性、5 倍以上のバイオフィルム除去能をもつ変異体の取得に成功した。また、これまでに directed evolution (DE) と machine-learning-guided DE について同一の条件下で比較した研究は報告されておらず、zero-shot generation を含めた手法ごとのタンパク質の進化の比較を行った (図 2)。その結果 MLDE が広い探索空間と高い成功率を収めることがわかった。



先行研究と比べて何がすごい？技術やアプローチのキモはどこ？

従来の MLDE では「スループットが低い系で DE がそもそも成立しない」「変異座位を数残基に限定」などの条件付き成功例が多かったが、TeleProt では **1ラウンド ≈ 1 万変異体** を扱えるマイクロ流体スクリーニングに合わせ、進化情報と実験情報を重み



付け統合する DNN / VAE、外挿リスクを抑える信頼半径（Hamming 距離）制約、目的距離分布を課した貪欲サブセット化などを組み込むことで広範な範囲の探索を最小限の実験コストで達成した。

どうやってこの手法/仮説の有効性を検証したのか

超高スループット選抜系の構築:

ドロップレットに *B. subtilis* 細胞を封入し、蛍光物質と消光物質をつけた DNA を用いて、DNA 切断に伴う蛍光増大で活性を読む実験系を構築。Generation2〜Generation4 で順次厳しい閾値を設定し、各ラウンド約 10000 変異体を並列評価。スループット性と解像度のバランスを加味した方法論により網羅性を維持しつつ、詳細な実験を行なっている。この実験系を用いて実際に作成した変異体と WT の活性および変異パターンの距離を調べた (図 3)。

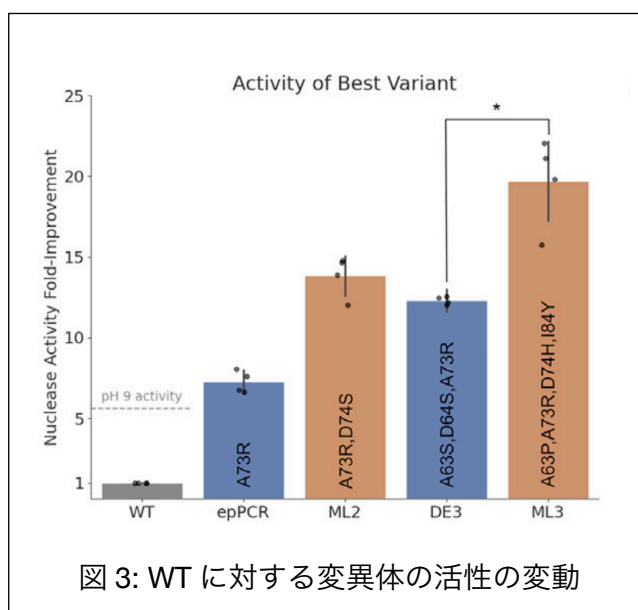


図 3: WT に対する変異体の活性の変動

また、各種法の hit rate を調べたことで、従来の DE よりも MLDE の方が高いヒット率であることを示した。また提案した MLDE の手法が広いパターンの変異を探索していることがわかった。

その他、議論した内容 (ネガティブコメントや limitation もあれば)

- 企業がこれまで色々と試してきた蓄積を論文としてまとめている印象を受けた。
- 今回は活性だけに着目しているが、生物に産生させることを考えた時にあまりにも多くの変異を入れると成長自体にも影響を与えかねない。なので、必ずしも変異がたくさん入っていることは工業的には嬉しい可能性はある。
- 進化によって得られた高い活性を持つ変異体について、なぜ活性が上がったのを理論的もしくは実験的に示すことができればさらに高いインパクトの論文になっただろう。
- 今回の実験の系は DNase の評価にしか使えない。より一般化した fitness の計測の系を作ることができればさらに色々できるのでは？