

TSUBAME 共同利用 令和 5 年度 学術利用 成果報告書

深層生成モデルによる人工タンパク質設計
De novo protein design with deep generative models

島村 徹平
Teppei Shimamura

東京医科歯科大学 難治疾患研究所
Medical Research Institute, Tokyo Medical and Dental University
<https://www.shimamlab.info/>

本研究では、深層生成モデルと分子動力学シミュレーションを用いて、特定の作用をもつタンパク質を *de novo* で設計する技術を開発する。特に、特定のタンパク質モジュールを阻害する人工タンパク質に焦点を当て、ロボット技術と実験検証を組み合わせることで、これらのタンパク質を自動で設計、構築、検証、学習するプラットフォームを開発する。最終的に、設計されたタンパク質を用いた新たながん治療法の開発を目指す。本年度は、プラットフォームのプロトタイプ開発、およびがんにおいて発現上昇が見られる代謝酵素である BCAT1 に対して阻害活性を有する候補タンパク質の設計を進めた。

In this study, we develop a technique for *de novo* design of proteins with specific functions using deep generative models and molecular dynamics simulations. Specifically, we focus on *de novo* proteins that inhibit specific protein modules and develop a platform which automatically designs, builds, tests, and learns these *de novo* proteins by combining robotics and experimental validation. Ultimately, we aim to develop new cancer treatments using the designed proteins. This fiscal year, we have made progress in developing a prototype of the platform and designing candidate proteins with inhibitory activity against BCAT1, a metabolic enzyme upregulated in cancer.

Keywords: タンパク質設計、深層生成モデル、分子動力学シミュレーション、ロボティクス、BCAT1

背景と目的

近年の急速な AI 技術の発展により、タンパク質の配列や構造を *de novo* 生成する深層学習モデルが注目を集めている。特に、与えられた条件にしたがってタンパク質を生成する構造ベースの生成モデルである RFdiffusion[1] や、アミノ酸配列を無条件に生成する配列ベースの生成モデルである EvoDiff[2] など、タンパク質生成における強力な生成モデルが数多く開発されている。しかしながら、このように生成されたタンパク質は機能や活性において既存のタンパク質に劣ることも多く、高機能・高活性なタンパク質を設計する技術の開発が急務である。

そこで本研究では、ロボット技術と実験検証を組み合わせることで、タンパク質を自動で設計、構築、検証、学習するプラットフォームを開発する。最終的に、設計されたタンパク質を用いた新たながん治療法の開発を目指す。

概要

本研究では、タンパク質を自動で設計 (Design)、構築 (Build)、検証 (Test)、学習 (Learn) する DBTL サイクルプラットフォームを開発する。個々の構成要素は以下からなる。

- ・設計: 深層生成モデルによるタンパク質生成
- ・構築: 大腸菌によるタンパク質合成
- ・検証: がん細胞株による細胞増殖抑制評価
- ・学習: ベイズ最適化を用いたフィードバック

設計・学習は AI が、構築・検証はロボットが行うことにより、DBTL サイクルを完全自動で回し、高機能・高活性なタンパク質を迅速に創出するための基盤技術を実装する。

さらに、実際にこの基盤技術を用いることで、がんにおいて発現上昇が見られる代謝酵素である BCAT1 に対して阻害活性を有する候補タンパク質の設計を行う。

結果および考察

設計: 標的タンパク質の PDB ファイルを入力として、RFDiffusion[1]による阻害タンパク質の構造生成、ProteinMPNN[3]による阻害タンパク質の配列生成、AlphaFold-Multimer[4]による標的タンパク質-阻害タンパク質複合体構造予測とフィルタリングを行うパイプラインを TSUBAME 上に実装した。また、フィルタリングされた阻害タンパク質配列を基に、コンビナトリアルアセンブリ法により大量の改変配列を生成する手法を実装した。

評価: *in vitro* 評価だけでなく *in silico* 評価も行えるように、GROMACS[5]を用いた分子動力学シミュレーション、および結合エネルギー計算を end-to-end で行えるパイプラインを実装した。

学習: 評価フェーズで得られた結果を設計にフィードバックするための多目的バッチベイズ最適化手法を BoTorch で実装した。代理モデルとして、阻害タンパク質配列を入力とし、評価値を推定する深層アンサンブルを PyTorch で実装した。

BCAT1 阻害タンパク質の設計: 生成 AI に対する概念実証として、BCAT1 に対する阻害タンパク質を生成 AI により設計し、*in vitro* 実験によりがん細胞増殖抑制能を確認した(図 1)。

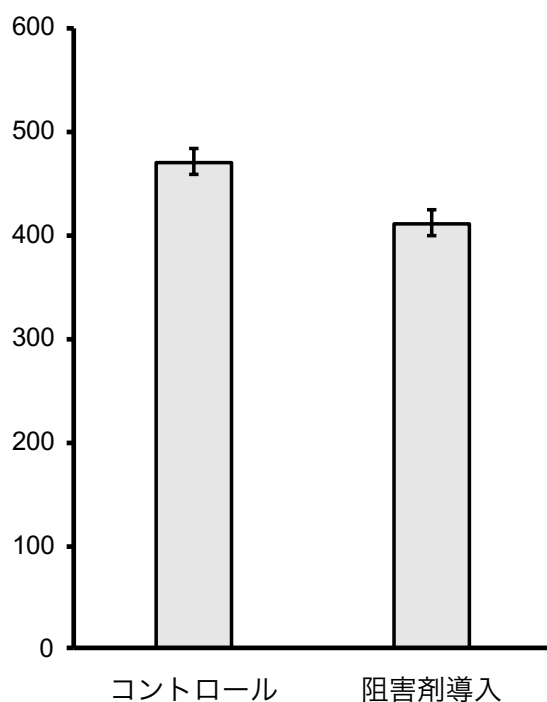


図 1: がん細胞抑制能の評価

まとめ、今後の課題

DBTL サイクルのうち、設計・*in silico* 評価・学習を TSUBAME 上に実装し、BCAT1 阻害タンパク質の設計を通して生成 AI に対する概念実証を行った。今後は DBTL サイクルを回すことにより BCAT1 阻害タンパク質の阻害活性を向上させるとともに、ロボットと AI のシームレスな接続を行うための基盤技術開発を目指す。

参考文献

- [1] Watson et al. De novo design of protein structure and function with RFDiffusion. Nature. 2023;620(7976):1089-1100.
- [2] Alamdari et al. Protein generation with evolutionary diffusion: sequence is all you need. bioRxiv. 2023:556673.
- [3] Dauparas et al. Robust deep learning-based protein sequence design using ProteinMPNN. Science. 2022;378(6615):49-56.
- [4] Evans et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv. 2021:463034.
- [5] Abraham et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1:19-25.