

2023 年度年次報告書

データ駆動・AI 駆動を中心としたデジタルトランスフォーメーションによる生命科学研究の革新

2023 年度採択研究代表者

浅井 潔

東京大学 大学院新領域創成科学研究科
教授

生物情報アーカイブを活用した深層生成モデルによる mRNA 最適設計技術

主たる共同研究者:

阿部 洋 (名古屋大学 理学研究科 教授)

齋藤 裕 (産業技術総合研究所 人工知能研究センター 主任研究員)

佐藤 健吾 (東京電機大学 システムデザイン工学部 教授)

研究成果の概要

拡散モデルによる配列生成に関して、RNA 言語モデルやその他の手法によって RNA を拡散モデルの連続空間に埋め込む方法を検討し、5' UTR 配列に適用した。形成されうる2次構造の自由エネルギーによって条件付けして本手法で RNA 配列を生成し、その分布を評価した。また、RNA2次構造予測 MXfold2 の改良を行い、chemical probing データからの学習を可能とした。

言語モデルによる配列生成に関して、生物種情報とタンパク質言語モデルによる潜在表現から mRNA 配列を生成する逆翻訳言語モデルのプロトタイプを開発し、動作を確認した。また、ribosome profiling データと AlphaFold2 予測構造の相関解析を行い、様々な生物種で翻訳速度と相関する構造特徴量を発見した[1]。さらに、ribosome profiling データの深層学習により、由来生物種での翻訳速度を模倣するように宿主生物種における mRNA 配列の設計手法を開発した。

mRNA の性能予測に関して、3種類モデルの開発を行った。利用可能な公開データの調査に基づき、由来の異なる複数の学習データを作成した。翻訳量予測モデルにおいては、深層学習を用いることにより既存の指標を上回る予測精度が得られることを見出した。分解耐性予測モデルにおいては、既存アルゴリズムの精査と検証を行うとともに、対照学習などを用いた改良を試みた。転写量予測モデルについてはシミュレーションデータによる学習を予備実験として実施した。

mRNA の配列合成と機能値測定課題においては mRNA の合成手法の確立とレポータータンパク質の選定を行った。PureCap analog の最適化と T7 promoter の改変により効率的な 5'Cap 化 mRNA の合成手法を確立した。また d2EGFP をレポーターとして利用することでハイスループットな機能値測定を行う態勢を構築し、多量のサンプルを評価する準備を整えた。

【代表的な原著論文情報】

- 1) Bian Bian, Toshitaka Kumagai, Yutaka Saito. VeloPro: A pipeline integrating Ribo-seq and AlphaFold deciphers association patterns between translation velocity and protein structure features. iMeta, 2(4):e148, 2023. doi: 10.1002/imt2.148.