

2024 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 (北里大学 未来工学部 教授)

佐藤 健吾 (東京科学大学 生命理工学院 教授)

研究成果の概要

いくつかの条件を組み合わせることで所望の機能値を持つ mRNA 配列の設計を可能とする拡散モデルを実装し、その性能を確認した。翻訳効率や自由エネルギーなどの条件では適切な配列が生成されるものの、安定性などの一部の条件では所望の機能値を持つ mRNA 配列を生成できないことが確認された。さらに、深層配列生成モデルの精度向上に配列性能予測モデルによるデータ拡張が有効であることを確認した。

逆翻訳言語モデルの開発に必要となる RNA 生成言語モデルについて、独自の事前学習モデルを開発した。既存モデルはノンコーディング RNA のデータで事前学習されているが、本研究では mRNA に特化した事前学習モデルを新たに開発した。本モデルを逆翻訳言語モデルに導入することにより、mRNA 配列生成の精度（天然配列との一致度）を向上することに成功した。また、酵母の全遺伝子の mRNA を大腸菌発現用にデザインする計算機実験を行ったところ、生成された配列は大腸菌でのコドン適合度が向上していることを確認した。これにより、逆翻訳言語モデルを用いて、生物種依存的に天然配列を模倣するような配列生成が行えることが示唆された。

mRNA の性能予測モデルの拡張および高度化を行った。RNA 分解耐性予測モデルについては、弱教師あり学習と半教師あり学習を組み合わせたパイプラインを構築し、既存のツールよりも高精度な予測が可能であることを見出した。翻訳量予測モデルに対しては、複数の ribosome profiling データを用いたマルチタスク学習を導入することにより予測精度が向上することを確認し、あわせて翻訳量予測に寄与する新たな特徴量も抽出した。

mRNA の配列合成と機能値測定課題においてはハイスループットな mRNA の調製手法を確立した。2D-HPLC の条件検討を行い、5'側の Cap 構造および 3'側の poly A を有する完全長の mRNA を精製する体制を構築した。これにより自動的に多量の mRNA を、副産物の影響を排除しつつ調製する手法を確立した。また GFP, NanoLuciferase ベースのレポーターについて評価を行い、ハイスループット機能値測定を行う体制を構築し、多量のサンプルを評価する準備を完了した。

【代表的な原著論文情報】

- 1) Jiyuan Yang, Kengo Sato, Martin Loza, Sung-Joon Park, Kenta Nakai. RNA Secondary Structure Prediction by Conducting Multi-Class Classifications. *Comput. Struct. Biotechnol. J.*, 27:1449-1459, 2025. doi: 10.1016/j.csbj. 2025.04.001.
- 2) Kosuke Fukuchi, Yuko Nakashima, Naoko Abe, Seigo Kimura, Fumitaka Hashiya, Yuichi Shichino, Yiwei Liu, Ryoko Ogisu, Satomi Sugiyama, Daisuke Kawaguchi, Masahito Inagaki, Zheyu Meng, Shiryu Kajihara, Mizuki Tada, Satoshi Uchida, Ting-Ting Li, Ramkrishna Maity, Tairin Kawasaki, Yasuaki Kimura, Shintaro Iwasaki, Hiroshi Abe. Internal cap-initiated translation for efficient protein production from circular mRNA. *Nat. Biotechnol.*, 2025, doi: 10.1038/s41587-025-02561-8.
- 3) Kazuteru Yamamura, Kiyoshi Asai, Junichi Iwakiri, Consistent features observed in structural probing data of eukaryotic RNAs, NAR Genomics and Bioinformatics, Volume 7, Issue 1, March 2025, lqaf001, doi: 10.1093/nargab/lqaf001.