

京都大学	博士（ 医 学 ）	氏名	角 俊 輔
論文題目	Deep generative design of RNA family sequences (RNA ファミリー配列の深層生成設計)		
(論文内容の要旨)			
過去数十年にわたり、RNA の機能性を活用することで多様な合成分子システムが発明され、分子生物学・医学・バイオマニュファクチャなど多くの分野に大きな影響を与えてきた。これら多くの実験研究では、リボレギュレーターやリボザイムなどの特定の機能性 RNA を設計するために、反復的な試行錯誤やランダム変異体探索が行われている。しかし、配列の多様性は長さの指数関数的に広がるため、実験的探索には限界がある。そこで、効率的に所望の配列を設計する計算機的プラットフォームが長らく求められていた。 そこで、本研究では、深層生成モデルを用いたデータ駆動的 RNA 設計手法の確立を目指した。深層生成モデルは、既存データの特徴を学習して疑似的なデータを生成することから、既存の RNA を模倣しながら、新規特性を有する RNA を設計することを目指した。加えて、機能性 RNA の多くは、RNA family という配列保存性と二次構造を共有する配列群を形成する。この配列保存性と二次構造は Covariance Model (CM) によってモデル化されることが知られる。そこで、深層生成モデルと CM を組み合わせることで、既存の特徴を精度良く再現しながら新規配列が設計できるのではないかと考えた。このような手法として、本研究では Variational Autoencoder (VAE) と CM を組み合わせた RfamGen を提案した。RfamGen は CM における配列表現を VAE により生成する深層生成モデルであり、特定の RNA family の配列を新規設計することが可能である。RfamGen は、潜在空間という情報の集約された低次元空間を有しており、これによって配列生成が制御可能のため、所望する特性の設計が可能であると考えられた。 本研究では、CM による配列の生成能への影響を検討するために、二次構造と配列保存性の要素をそれぞれ欠損させたモデルを作成した。各モデルが生成した配列の RNA family としての尤度を比較したところ、RfamGen の生成配列が最も高い尤度を示し、RfamGen の高い生成能が示された。 また、実験検証として、100~600 程配列のみ報告される 3 つの RNA family を対象に RfamGen を適用し、生成配列の機能検討を行った。すると、生成された配列全てにおいて機能が認められ、少数データからでもあっても RfamGen は安定的に配列を生成可能であることが示された。 最後に、配列の設計成功率をより詳細に検討するために、自己切断 ribozyme の <i>glmS</i> ribozyme を用いて RfamGen を学習し、生成された 1000 配列に対して大規模並列アッセイを実施することで機能性の網羅検討を行った。このために、新規に大規模並列アッセイとそのデータ解析パイプラインを構築した。比較対象としてデータベースに登録されるほぼ全ての 800 配列と既知活性変異体 100 配列を加えた RNA ライブラリを構築し、全ての配列に対して <i>glmS</i>ribozyme の機能評価を行った。その結果、ほぼ全ての生成配列において自己切断活性が認められた。また、天然配列と生成配列の活性強度を比較したところ、生成配列が優位に高い活性を有することが明らかになり、RfamGen が配列活性の向上に有用であることが示唆された。 以上の結果から、RfamGen は、高い生成能を有する RNA family 配列の生成モデルであり、 RNA 設計における新規プラットフォームとなりうる。これにより、RNA 工学分野が発展し、医学や産業への RNA 応用が促進されると期待される。			

(論文審査の結果の要旨) RNA 工学とは、機能性 RNA を改変・設計することにより、医学生物学や生命工学に応用する試みである。しかし、従来まで計算機的設計手法に乏しく、機能性 RNA 設計には技術的な難易度が高い。 本研究では、この課題に対して、深層生成モデルを用いた RNA 設計というアプローチを試みた。このために、機能性 RNA の分類である “RNA ファミリー” に着目し、特定の RNA ファミリーの配列を生成する深層生成モデル RfamGen (RNA family sequence generator) を開発した。 RfamGen は、共分散モデルというアライメントと二次構造を考慮したモデルと深層生成モデルを統合することにより、RNA ファミリー配列設計のデータ効率性が高いことが明らかになった。また、RfamGen の潜在空間の可視化により、配列情報を集約し RNA 設計に有用であることが示された。さらに、自己切断型 RNA 酵素の RNA ファミリーを対象に RfamGen を適用し、機能性検証実験を行うことで RfamGen のデータ効率性と生成配列の機能性が示された。最後に、大規模並列実験による生成配列の機能評価が実施され、RfamGen の高い設計成功率が示された。 以上の研究は、機能性 RNA の計算機的設計基盤を提供するものであり、RNA 工学と医学生物学への応用研究全般に寄与するところが多い。 したがって、本論文は博士(医学)の学位論文として価値あるものと認める。 なお、本学位授与申請者は、令和6年1月30日実施の論文内容とそれに関連した試問を受け、合格と認められたものである。
