

(続紙 1)

京都大学	博士 (情報 学)	氏名	Unyanee Poolsap
論文題目	Computational Methods for Predictions of RNA Pseudoknotted Secondary Structures and RNA-RNA Interactions シュードノットつきRNA 2 次構造予測とRNA間相互作用予測に対する計算手法		
(論文内容の要旨)			
本論文はRNA配列データからのRNA二次構造およびRNA間相互作用の予測方式について述べられており、5 章から構成されている。 第 1 章ではRNA分子の重要性に関する議論を中心に、本論文で対象とする問題の生物学的背景について述べている。 第 2 章では本論文で必要となるRNA分子の二次構造および相互作用についての生物学的基礎事項、および、本論文で用いる情報技術である整数計画法および動的計画法の概要について説明している。 第 3 章ではRNA配列からシュードノット付きのRNA二次構造を整数計画法を用いて予測する方法を提案している。提案手法ではシュードノットつきRNA二次構造をシュードノットなしのRNA二次構造のペアとして表現し、一つの塩基は高々一つの塩基とのみ対を形成する、シュードノットなし構造においては対を表す辺が交差しない、塩基対を形成する場合には少なくとも二個の塩基が連続しなくてはならない、などの必要な制約を整数制約のもとでの線形不等式として表現し、RNA二次構造予測問題をこれらの制約のもとで連続する塩基対間の結合エネルギーの和を最小化する整数計画問題として定式化している。そして、この定式化のもとで整数計画問題に対する既存のソルバーを適用することにより予測を行っている。データベースに登録されている既知のRNA二次構造データを用いた既存手法との計算機実験による比較の結果、少なくとも塩基数が 70 個以下の短いRNA配列に対しては提案手法は既存手法と同等以上の予測精度を持つことを示している。また、提案手法はRNA二次構造がシュードノットを持つ場合には、ほとんどシュードノットを出力することも示している。さらに、長い配列に対する予測精度が既存手法より優れていない理由やその方策についても議論している。 第 4 章ではRNA間相互作用を動的計画法を用いて予測する方法を提案している。既存手法と比較し大幅に計算時間を改善することを主目的として、既存の方法論とは異なり、一方のRNAの結合部位が既知の場合に、主対象となるもう一方のRNAの結合部位、および、その際の内部二次構造を予測する問題として予測問題を定式化しなしている。この制約のもとで高精度の予測を行うために、まず既存手法を用いて結合部位の候補を予測し、次に、統計的手法により得られたスコア関数と動的計画法を組み合わせることにより、結合部位も含めて期待精度が最大となる最適な二次構造を計算する方法を提案している。そして結合部位の個数やサイズが定数で抑えられるという妥当な仮定のもとで、この手法が入力配列長の 3 乗のオーダーの計算時間で動作することを示している。提案手法を評価するために、結合二次構造が既知のRNA間相互作用データを用いて既存手法との計算機実験により比較を行い、比較的長いサイズのRNAについては提案手法が既存手法より高い予測精度を持つこと、および、計算時間についても提案手法が既存手法より同等以上に高速であることを示している。さらに、一方のRNA配列を与えた際にゲノム中の多くの配列との結合二次構造を計算することにより、新たな結合RNA対を検出するという問題についても計算機実験を行い、その結果として統計的に有意なRNA対を発見できる可能性があることを示している。 第 5 章は結論であり、本研究をまとめるとともに、提案手法の拡張の可能性や今後の課題について述べている。			

(論文審査の結果の要旨)

本論文は、重要な生体内高分子であるRNAの構造に関して、整数計画法を用いたシュードノットつきRNA二次構造の予測方式、および、動的計画法を用いたRNA間相互作用部位の予測方式について述べたものであり、得られた成果は以下のとおりである。

(1) RNAの生体内および細胞内で果たす役割についての重要性が近年再認識され、それに伴い、RNA二次構造と呼ばれる結合塩基対のなす構造の予測について様々な計算手法が提案されるようになってきた。その一つとしてシュードノットとよばれる、複雑に交差した構造を含む二次構造の予測が多く研究されるようになったが、シュードノットなしの場合と比較して、計算量が大きく、かつ、アルゴリズムも複雑になるという問題があった。さらに非常に一般性の高いシュードノットを扱う場合には最適構造の予測がNP困難になるということも知られていた。そこで本論文では、RNA配列データからのシュードノットつきRNA二次構造の予測問題を整数計画問題として定式化し、そこに既存のソルバーを適用して最適構造を計算することにより予測を行う方式を開発した。具体的には、シュードノットつきの二次構造が満たすべき条件を整数制約のもとでの線形不等式で与え、その制約下で連続する塩基対のエネルギーの和を最小化する問題として定式化した。既存手法との計算機実験による比較の結果、提案手法は長さ70塩基以下の短いRNA配列に対しては既存手法と同程度以上の計算速度で同程度以上の予測精度が得られることがわかり、その有用性が確認できた。さらに提案手法は様々な制約を比較的容易に導入できることから、拡張性の高い方法であることも示唆された。

(2) 近年の生物学の進展によりRNAは生体内で様々な制御に関与していることが明らかになりつつある。RNA間相互作用もその重要な制御機構の一つであり、その予測に関して多くの研究が行われるようになってきた。特に近年、結合二次構造の予測手法がいくつか提案されてきたが、既存手法には単純な構造しか考慮されていないか、もしくは、膨大な計算時間がかかるかのいずれかの問題点があった。そこで、2個のRNA配列データから直接結合二次構造を予測するのではなく、一方のRNA配列のアクセシビリティを計算し、その情報を利用して効率的にもう一方のRNAの結合部位および結合二次構造を予測する手法を開発した。結合部位に関する情報を動的計画法に巧妙に埋め込むことにより、結合部位のサイズおよび個数が定数で抑えられるとの妥当な仮定のもとで、入力配列長の3乗のオーダーの計算時間で動作する予測手法を開発した。そして最新の統計的手法に基づくスコア関数と提案手法を組み合わせ、既存の結合二次構造予測手法との計算機実験による比較を行った結果、提案手法は既存手法と同等以上に高速であり、かつ、比較的長い配列データについては、より高精度の予測が行えることが示された。また、多数の配列との結合部位予測に適用することにより、提案手法が新たな相互作用RNA対の予測にも有効であることを示唆する結果も得た。

以上、本論文ではバイオインフォマティクスにおける重要な研究テーマである、RNA配列データからのRNA二次構造予測、および、RNA間相互作用予測の両者に関して、新規な手法を提案するとともに、実際のRNAデータを用いた計算機実験により、それらの有効性を示している。提案手法のいずれもが新規性、有用性、拡張性の高いものであり、当該分野の発展のために十分な寄与をしている。よって、本論文は博士(情報学)の学位論文として価値あるものと認める。

また、平成２３年６月１６日論文内容とそれに関連した口頭試問を行った結果合格と認めた。