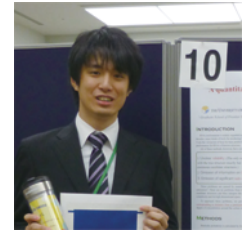


RNA 二次構造予測における 本質的曖昧さを克服する新手法の開発

森 遼太

東京大学大学院 新領域創成科学研究科
情報生命科学専攻 浅井研究室



生体内で転写されたRNA分子は熱力学的に安定な高次構造を構成しますが、この構造は多くのRNAの機能と密接に関わっていることが知られています。高次構造のうち最も低次のカテゴリとして、分子内の塩基対形成のみを考慮した二次構造と呼ばれるものがあります。二次構造レベルでも多くの生命科学的知見が得られるということに加え、より高次の構造を確認する手法は未だ確立していないということから、これまで実験的手法と理論的手法の双方でRNA二次構造を予測しようという試みが行われてきました。

しかしながら、熱力学的なゆらぎを考慮すると、特定の構造自体が生体内に存在する確率は一般に極めて低く、例えば凡そ300塩基程度の分子の場合、たとえ最も安定な構造を選択したとしてもその存在確率は 10^{-10} 以下のような値になってしまいます。従って、多くの先行研究のように、ある一つの構造を予測するような推定手法というものは、得られた構造をどの程度信頼してよいか分からない上、ゆらぎ方や準最適構造に関する情報が欠落するなど、多くの生命科学的知見の見落としや誤謬に繋がるのが考えられます。

そのような中私たちは予測したある構造に対して、その構造との類似度に応じたどの程度の割合の構造が集中しているかを計算することができれば、既存手法が抱えていた多くの問題点を解決することが出来るのではないかと考えました。丁度リボスイッチと呼ばれるRNAの発見を目的に同様の分布を計算する手法が近年提案されましたが、これは実際には計算量的な問題を抱えていました。そこで私たちは離散フーリエ変換を応用することで劇的に計算量を抑え、大規模なRNAに対して現実的な時間でこの分布を得るアルゴリズムを考案しました。また、この計算量的優位性を活かして本手法を複数の異なる構造それぞれとの類似度に応じた存在確率が計算できるよう拡張し、構造間の比較も可能にしました。

本手法の出力例を[図1]及び[図2]に示します。[図1]は、あるtRNAに関してその重心構造(熱力学的ゆらぎの中心

のような構造)との類似度に応じた確率の分布を表しています。重心構造自体の存在確率は1%にも満たないのですが確かに重心構造周辺にかなりの存在確率が集中していることが分かります。しかしながら同時に、重心構造近傍のみでは説明できないような構造も無視できない割合で存在しているように見て取ることもできます。実際、tRNAの二次構造として有名なクローバーリーフ構造は、図中に赤い矢印で示した付近に存在しており、既存の構造予測手法では気付けない重要な構造の存在を発見することが出来た一つの例と言えます。また、[図2]は同じtRNAについて重心構造とクローバーリーフ構造の双方からの類似度に応じた構造の存在確率分布です。[図2]から、両者の間には高いポテンシャル障壁があり、また両者の他に目立った第三の準最適構造は無さそうであるということが分かります。

現在は以上の手法を利用し、自動的にRNAの分布様態を要約するソフトウェアを作成しながら、網羅的な解析による新たな生命科学的知見の獲得を目指しています。

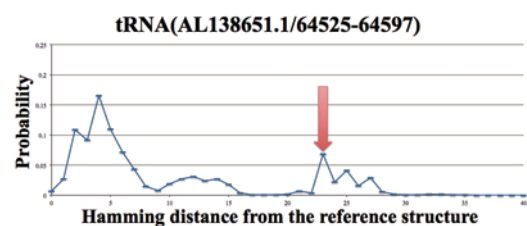


図 1

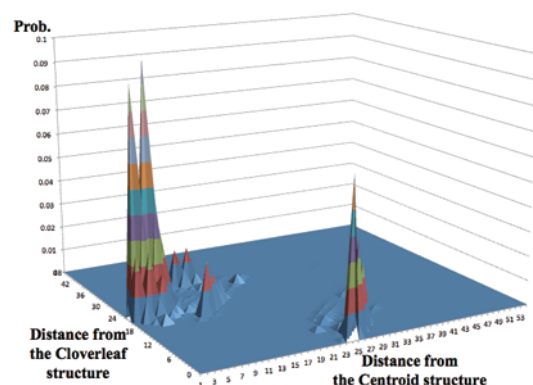


図 2