

平成 20 年 12 月 16 日

東京大学大学院新領域創成科学研究科

「RNA の二次構造予測で世界最高精度のソフトウェアを開発」

1. 解禁日時: 平成 20 年 12 月 18 日午前 0 時 0 分
2. 発表者: 浅井 潔(大学院新領域創成科学研究科 情報生命科学専攻 教授)
3. 発表雑誌: Bioinformatics (Oxford University Press、英国)

4. 発表概要:

東京大学、(独)産業技術総合研究所、みずほ情報総研(株)、(社)バイオ産業情報化コンソーシアムの共同研究チームは、生体内の重要な高分子である RNA の二次構造の予測に関して、予測精度を最大化する独自理論に基づき、現時点で世界最高精度のソフトウェアを開発した。

5. 問い合わせ先

〒277-8561 千葉県柏市柏の葉 5-1-5 基盤棟 CB04

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

教授 浅井 潔

解禁日時：平成 20 年 12 月 18 日午前 0 時

RNA の二次構造予測で世界最高精度のソフトウェアを開発

平成 20 年 12 月 16 日

東京大学大学院新領域創成科学研究科の浅井潔教授、みずほ情報総研(株)の浜田道昭氏、(独)産業技術総合研究所生命情報科学研究センターの光山統泰研究チーム長、木立尚孝特別研究員、(社)バイオ産業情報化コンソーシアムの佐藤健吾研究員の共同研究チームは、生体内の重要な高分子である RNA の二次構造予測に関して、予測精度の期待値を最大化する独自理論に基づき、現時点で世界最高精度のソフトウェア、「CentroidFold」を開発した。

RNA は遺伝子 DNA からタンパク質が合成される際の中継ぎとしての役割が古くから知られているが、近年そのような役割をしない RNA が遺伝子の発現制御や細胞のガン化などで重要な役割を果たしていることが分かってきた。このような RNA を 機能性 RNA というが、機能を発現するには長い 1 本鎖である RNA 分子が部分的に 2 本鎖を形成して二次構造と呼ばれる特異的な構造をつくる。

1 本の長い RNA 鎖がとる塩基対(AとU、GとC)の組み合わせは膨大な数になるが、RNA を安定な状態にする塩基対の組み合わせは限られており、その数は1つではない。従来の予測では最も安定(エネルギーが最小)な二次構造を正しい二次構造として求める方法がとられてきたが、最も安定な二次構造が必ずしも正しい二次構造であるとは限らない場合が多々あることが経験的にわかってきた。

今回、開発した RNA 二次構造予測手法は、同一のエネルギーモデルを用いた場合には 予測精度の期待値が最大になることを理論的に証明し、ベンチマークテストでも世界最高の精度で二次構造を予測できることを実証した。二次構造予測の技術は、生物学実験では欠かせない DNA 増幅(PCR) に用いるプライマーの設計や、近年医薬品としての用途を開拓しつつある RNA 干渉 に用いる siRNA の設計、さらにはマイクロアレイに用いるプローブの設計など、バイオテクノロジー産業に広く影響を与える基盤技術である。今後、RNA 研究における二次構造予測の標準的なソフトウェアとして使われると共に、機能性 RNA の発見や、機能解明に貢献することが期待される。「CentroidFold」は、平成 20 年 12 月 18 日から、フリーソフトウェアとして無償で提供されると共に、インターネットから直接二次構造予測を行うサービスが提供される(<http://www.ncRNA.org/>)。

本研究は、独立行政法人 新エネルギー・技術総合開発機構(NEDO)の「機能性 RNA プロジェクト」によって行われた。

また、本研究は、近くイギリスの論文誌 Bioinformatics 誌に掲載される予定。

RNA 二次構造予測の重要性

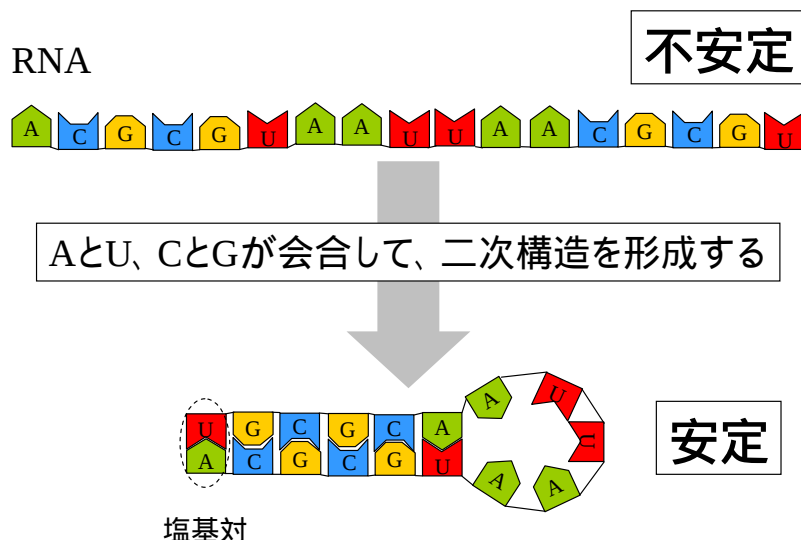
従来 RNA は DNA 配列を元にタンパク質を合成するための中間物質と考えられていたが、近年 RNA が生体内で様々な機能を担っていることが明らかになり、生命の機能を解明するための重要な物質として位置づけられてきた。このような RNA を機能性 RNA というが、機能性 RNA を利用した医薬品の開発も進んでいる。

RNA が機能を発現するのは 1 本鎖の長い分子である RNA 分子が、部分的に 2 本鎖を形成して特異的な構造を形成することに由来する。2 本鎖を形成するのは、RNA を構成する 4 種の塩基 (A、G、C、U) のうち、A と U、G と C がお互いに引き合うためである。1 本鎖 RNA における塩基対形成の様子を二次構造とよぶ。(立体構造は三次構造ともよぶ)。

RNA の二次構造を解明することは RNA の機能解明の重要な手がかりとなる。しかし、二次構造の実験における観測は大きな手間とコストが必要となるため、計算機による二次構造予測が RNA 研究における重要な手段となっている。

二次構造予測の技術は、生物学実験では欠かせない DNA 増幅 (PCR) に用いるプライマーの設計や、近年医薬品としての用途を開拓しつつある RNA 干渉に用いる siRNA の設計、さらにはマイクロアレイに用いるプローブの設計など、バイオテクノロジー産業に広く影響を与える基盤技術である。

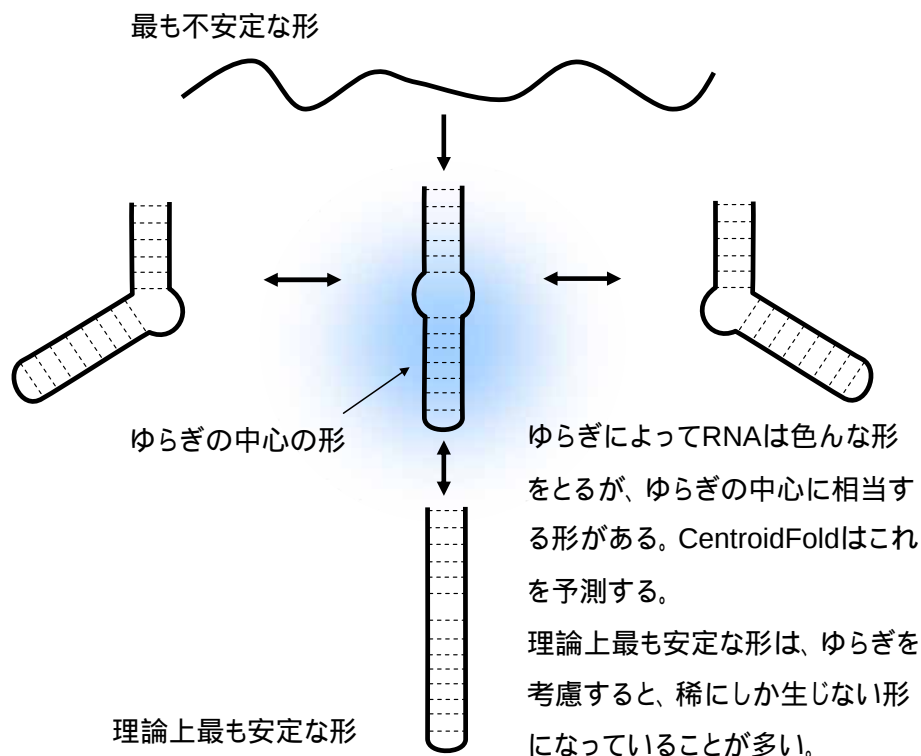
二次構造予測がより正確になることによって、より高い精度の siRNA 設計が可能になったり、生体内における RNA 分子の機能解明をより正確に推定したりすることが可能になるなど、RNA 医薬品開発のための技術水準を一段高めることになる。



背景となる理論と成果

1本の長いRNA鎖がとる塩基対(AとU、GとC)の組み合わせは膨大な数になるが、RNAを安定な状態にする塩基対の組み合わせは限られており、その数は1つではない。従来の予測ではエネルギー的に最も安定な二次構造を正しい二次構造として求める方法がとられてきた。しかし、細胞内の物質は1つの構造に固定されているのではなく、熱のエネルギーによって常に形を変化させながらゆらいでおり、RNAもまたゆらいでいると考えられる。そのため、個々の二次構造のとりやすさ(確率)は、理論上最も安定な二次構造であってもきわめて小さく、このことが、従来の予測方法で正しい二次構造を予測できない原因ではないかということが、最近の研究からわかってきた。

共同研究チームでは、エネルギーモデルから計算される二次構造の確率に基づいた予測精度の期待値が理論的に最大となる予測手法を発見し、ゆらぎの影響を受けにくい二次構造予測技術の開発に成功した。この技術に基づき、単独のRNA配列の二次構造予測とRNA配列群の共通二次構造予測を行うソフトウェア「CentroidFold」を開発した。「CentroidFold」は、RNAの二次構造予測の精度を評価するために広く用いられているベンチマークデータセット(構造がわかっているRNA)を用いた計算機予測実験でも、従来のRNA二次構造予測手法より高精度であることが実証された。



【用語解説】

RNA

リボ核酸の略称。生命活動に不可欠な物質の一種で、遺伝情報を読み取る役割を担っている。遺伝情報は DNA から RNA に写し取られタンパク質が合成されるが、そのため RNA は中間的な存在と考えられてきた。

機能性 RNA

タンパク質に翻訳されることなく、自身が生体内で活性を持ち、転写 / 翻訳制御などの機能を有する RNA 全般を指す。機能性 RNA の機能と二次構造(さらには、立体構造)との間には関連がある場合が多いことが知られている。

予測精度の期待値

RNA の二次構造はそのエネルギーが低いほどその構造をとりやすい性質がある。二次構造ごとに熱力学的に計算された構造のとりやすさ(確率)から、二次構造予測の精度の確率的な期待値を求めることができる。

PCR

細胞抽出液などの試料に含まれる特定の DNA を増幅するのに広く用いられる実験方法。標的 DNA の選別には特定の塩基配列を持ったプライマーとよばれる低分子の DNA 配列が用いられる。プライマーは標的 DNA に特異的に結合する配列である必要がある。標的 DNA(RNA)が二次構造を形成する可能性があると、正常な PCR 反応を妨げてしまう可能性があるため、プライマー設計には標的 DNA(RNA)の二次構造を考慮する場合がある。

RNA 干渉(RNA interference:RNAi)

siRNA を細胞へ導入することにより、相補的な塩基配列をもつメッセンジャーRNA (mRNA) が分解され、遺伝子の発現が抑制される現象。これまで治療が困難であったがんや遺伝子疾患、感染症など、さまざまな疾患の治療を可能とする医薬品開発技術として期待されている。

siRNA (small interfering RNA)

RNA 干渉において、標的遺伝子の mRNA と相補的に結合する 20 ~ 22 塩基程度の短い RNA 分子。siRNA が mRNA に相補的に結合することによって、当該 mRNA の翻訳が阻害されることで、遺伝子の発現を抑制することができる。標的 mRNA 配列が二次構造を形成する可能性があると、siRNA が効果を示さない場合がある。従って、siRNA 設計には標的 mRNA の二次構造を考慮する場合がある。