

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成 25 年 5 月 30 日現在

機関番号：82626

研究種目：若手研究（B）

研究期間：2011～2012

課題番号：23700359

研究課題名（和文） マルチスケールシミュレーションによる核酸立体構造予測

研究課題名（英文） Tertiary structure prediction by multi-scale simulation

研究代表者

亀田 倫史（TOMOSHI KAMEDA）

独立行政法人産業技術総合研究所・生命情報工学研究センター・主任研究員

研究者番号：40415774

研究成果の概要（和文）：

RNA1 塩基を 3 つの球（P:リン、S:糖鎖、B:塩基）で表現する粗視化モデルを構築した。このモデルを用いて分子動力学シミュレーションを行ったところ、初期構造に依らず、天然状態における立体構造に正しく巻き戻った。

研究成果の概要（英文）：

We developed the coarse-grained model of RNA molecule. Our model consists of 3 type beads (P: Phosphate, S: sugar, B: Base). Using this model, we carried out molecular dynamics simulation of RNA, which predicted native tertiary structure correctly.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
交付決定額	3,600,000	1,080,000	4,680,000

研究分野：

科研費の分科・細目：

キーワード：蛋白質、RNA、立体構造予測、分子動力学（MD）

1. 研究開始当初の背景

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。DNA とは異なり、RNA は容易に折れ曲がり、RNA 1 分子内の塩基同士が結合し、特異的な立体構造を形成することができる。RNA が単体でも機能を発現することができるのは、RNA がこのような特異的な立体構造形成能をもつためである。つまり、RNA 立体構造を理解することは、RNA の生体機能を理解する上で重要であり、RNA の塩基配列からその立体構造を決定・予測することができれば、理学的な観点のみならず、工学、薬学に対しても、大きな影響を与える技術となることは間違いない。

しかし、実験による構造決定は手間とコス

トの面で現実的ではなく、計算機による構造予測が RNA 研究における重要な手段となっているが、現在行われている構造予測は、ほとんどが二次構造予測（図 1 左）であり、立体構造予測はほとんど行われていない。二次構造の予測で十分な場合もあるが、上で述べた RNA アプタマーなど、RNA が蛋白質と直接結合し相互作用することを考慮すると、二次構造を明らかにするだけでは、不十分であると言わざるを得ない。また、現在の二次構造予測は pseudoknot などの特殊な構造を予測することが困難であるなど、精度の面からみても、問題がある（図 1 左）。

一方で、全原子モデルによる立体構造予測も現実的ではない（図 1 右）。全原子モデルの分解能は高く、立体構造を詳細に記述することが可能であるが、原子数が多いために非常に長時間の計算を必要とする。しかも、核酸の力学的性質を記述する力場の性能が悪

いために、通常の分子動力学 (MD) シミュレーションを行っても、ランダムな構造から計算を始めて、正解構造を予測することはほとんど不可能である。


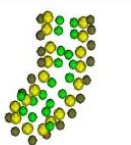
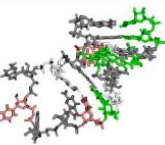
			
表現	二次構造	粗視化モデル 1塩基を3球(リン、糖、塩基)で表現	全原子モデル
計算速度	○数秒	△数分	×数日以上
制限	×Pseudo knotなど特殊な構造は予測不能	○特になし	○特になし
分解能	×低	△中間	○高
予測精度	○高(配列が短い場合)	○高(テスト計算中)	×悪い(全然大メ)

図1: RNAを表現する様々なモデル

2. 研究の目的

本研究では、RNAの立体構造予測を、粗視化モデル(図1中)と全原子モデル(図1右)とを組み合わせたマルチスケールシミュレーションを用いて行う。特に、既存の二次構造予測では予測が不可能な、Pseudoknotなどの特殊な構造の予測を目指す。

3. 研究の方法

(1) まず、粗視化モデルを用いた分子動力学計算によって、粗いレベルでの立体構造予測を極めて高速に行う。

(2) 次に、粗視化モデルの予測結果を距離に関する拘束条件として表現する。

(3) 最後に、この拘束条件と全原子モデルを組み合わせた分子動力学計算を行い、高分解能の立体構造予測を高速に行う。

特に、粗視化モデルを用いたMD計算によって、RNAの大まかな骨格(リン、糖鎖、塩基)が予測できるので、それを元にして、全原子モデルを作成するのは容易だと考えられる。そこで、実質的に困難な問題となるのは粗視化モデルの構築部分である(1)である。

4. 研究成果

H23年度は、RNAを表現する粗視化モデルの構築を行った。具体的には、球間の距離・角度・ねじれ角を拘束するばね、静電相互作用、vdW相互作用に関するパラメーターの最適化を行い、一定の成果を上げた(図2、3)。

例として PDBcode:2F88 について行った計算結果について述べる。のびきった構造を初期構造として、分子動力学(MD)計算を行ったところ、正しく正解構造(図2)に巻き戻った。また、計算中に生じる立体構造を、主成分分析法(PCA法)を用いて、クラスター分析したところ(図3)、最も多く出現した立体構造(第一クラスター)が正解構造であることがわかり、計算によって得られた構造が、偶然正解構造と一致したわけではないこ

とが示された。

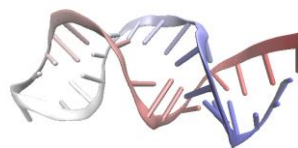


図2: RNA立体構造(塩基配列 GAGCCGUGUGCGAUGAAAGUCGCAAGCACGGUUC、PDBcode:2F88)

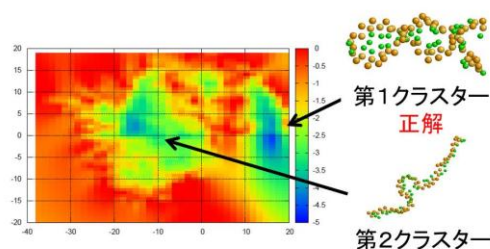


図3: 粗視化モデルを用いたMD計算例(PDBcode:2F88)

また、研究期間中に、RNA立体構造予測に関するコンテストを行った、との報告がなされたため(J. A. Cruz et al; RNA-Puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction; 2012 RNA) このコンテストで出題された問題についても、予測計算を行ったところ、形成するペアが一致するなど、良い結果を得た。(図4)

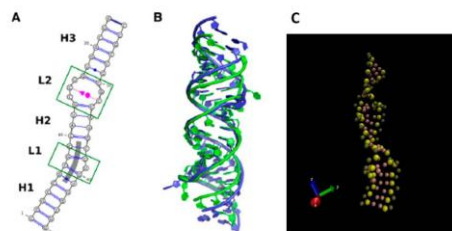


図4: RNA-Puzzlesで出題された問題(A)、コンテストにおける最もよい予測結果(B、緑が正解、青が予測)、本研究による予測(C)

この内容で、論文を投稿する予定であったが、他研究者による RNA 立体構造予測を行う研究がいくつか発表されたため、(例: Irina Tuszynska and Janusz M Bujnicki; 2011, BMC Bioinfo)、H24 年度では、現在のモデルの性能をさらに向上させてから発表をすることにした。具体的には、当初の予定よりもさらに多数の RNA に対して (100 個程度) 予測を行い成功させることを目標とした。そのために、H23 年度とは異なったアプローチで粗視化モデルの最適化を行った。というのは、PDB に登録されている RNA 単体からなる立体構造を多数集め、その中から配列の似た構造を排除し (配列同一度 $\leq 30\%$)、分解能が 3.0 Å 以下に絞った立体構造からなる統計データベースを作成し、そこから結合長などの平均値・分散を抽出し、モデル構築を試みた。また、RNA 単体の立体構造を解明するのは難しいことが知られており、このような基準でデータベースを作成したところ、高々 125 個の構造しか得ることができなかつたため、非常に長い RNA 配列で構成されているリボソームで、構造既知のものを用いて、同様のアプローチで統計データベースを作成し、比較したところ、単体 RNA の集合から得た統計値と、リボソームから得られた値とはほぼ等しいことがわかった。

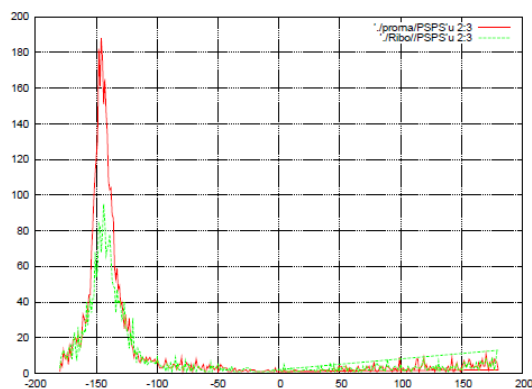


図 5 : 統計データベースから得られた、P-S-P-S 原子間の二面角の分布

赤は RNA 単体からなるデータベース。緑はリボソームからなるデータベース。

さらに、世界一の予測精度を持つ RNA 二次構造予測ソフト CentroidFold (東大・浜田道昭准教授・浅井潔教授) で用いられている、塩基対形成確率 (ContraFold モデルなど) を相互作用エネルギーとしてモデルに取り込み、立体構造予測を行った。新モデルは一定の成果は上げてはいるが、多数 (~100 個) の RNA に対して成功するレベルまでは到達していないため、現在改良を続けている。

また、上で述べた粗視化モデルの改良研究の過程で、各塩基に対して予測した塩基対ペア形成確率をその塩基について足し合わせ

た値が低ければ、その RNA 塩基が他の分子 (蛋白質) と相互作用している場合が統計的に有意に多いことを発見した。(図 6、7) つまり、ContraFold などの塩基対形成確率モデルは、RNA 単体の構造予測のみならず、RNA-蛋白質複合体の構造予測にも用いることができることを示した。また、様々なエネルギーモデルを用いても、結果はほぼ同等であることもわかった。(図 6、7) 本内容に関して論文を執筆し、現在投稿中である (J. Iwakiri, T. Kameda, K. Asai and M. Hamada; *submitted*).

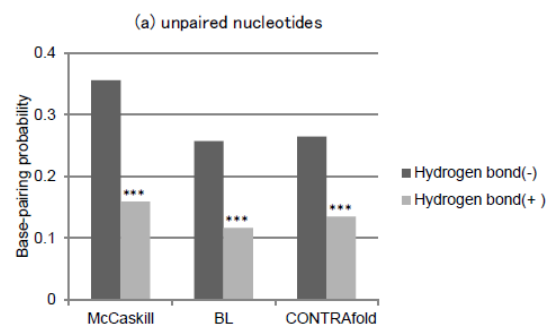


図 6 : 塩基対を形成しておらず、かつ、タンパク質と水素結合している (していない) RNA 塩基に対して、エネルギーモデルを用いて、塩基対形成する確率を計算し比較した。McCaskill、BL、CONTRAFold は、塩基対形成確率を計算する際に用いたモデル。

タンパク質と水素結合している塩基の方が、塩基対確率が有意に小さい。

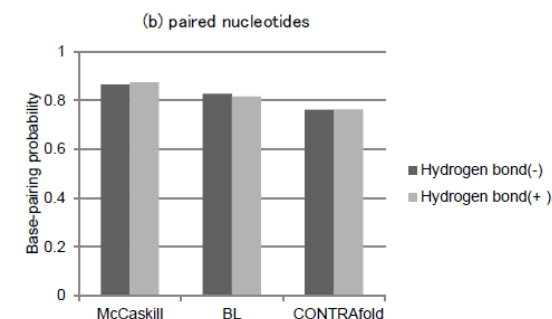


図 7 : 塩基対を形成しており、かつ、タンパク質と水素結合している (していない) RNA 塩基に対して、エネルギーモデルを用いて、塩基対形成する確率を計算し比較した。McCaskill、BL、CONTRAFold は、塩基対形成確率を計算する際に用いたモデル。

タンパク質との水素結合の有無にかかわらず、塩基対確率の値はほぼ同じ。

将来への展望としては、本章の後半で述べたように、RNA 単体についてのみならず、RNA-蛋白質、RNA-RNA の相互作用予測や複合体の立体構造予測に向けた知見も集まりつつある。そこで、本研究で行ったような MD シミュレーションのみに基づく予測法だけでな

く、ドッキングシミュレーションや、情報論的アプローチとシミュレーションを融合させた手法を用いれば、複合体構造予測などのより困難な問題を解くことができるかもしれない。実際、このような観点から蛋白質-RNA 相互作用を研究する、基盤研究 A「RNA・タンパク質相互作用の網羅的予測と検証」を H25 年度より開始したところである。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔学会発表〕(計 1 件)

亀田倫史、粗視化モデル分子動力学法を用いた RNA 立体構造予測、日本生物物理学会、2012 年 9 月 24 日、名古屋大学 (愛知県)

6. 研究組織

(1) 研究代表者

亀田 倫史 (TOMOSHI KAMEDA)

独立行政法人産業技術総合研究所・生命情報工学研究センター・主任研究員

研究者番号：40415774