



more info, slides, contact

# RNA 二次構造のループプロファイルと SHAPE 反応性の関係解析 — 構造予測とエネルギーパラメータ推定に向けて — Relationship Between RNA Loop Context Profiles and SHAPE Reactivity — A Step Toward Shape-Guided Folding and Energy Parameter Estimation —

Takumi Otagaki, Goro Terai,  
Junichi Iwakiri, Kazuteru Yamamura, Kiyoshi Asai  
University of Tokyo

LinkedIn:  
Takumi Otagaki

## Introduction

### Background:

- ◆ RNA の二次構造は、機能を支える重要な要素 (Fig1)。
- ◆ しかし既知の RNA 立体構造はタンパク質のわずか 3%にとどまる。(実験による決定は高コスト)
- ◆ SHAPE 実験は各塩基の reactivity (柔らかさの指標) を低成本かつ高スループットに測定可能。

### Challenge: Energy parameters for modified RNAs

- ◆ RNAには140以上の塩基修飾があり、それぞれ異なる安定性寄与 (エネルギーパラメータ) がある。
- ◆ 修飾塩基のエネルギーパラメータは、正確な構造予測に不可欠だが、効率的な推定法は未確立。
- ◆ SHAPEのような高スループット実験データを用いた計算的パラメータ推定が期待されている。(Fig2)

### Prior work & limitations

- ◆ 先行研究：ループの種類やループ内での位置と、SHAPE reactivity の相関が報告され、その情報を利用した二次構造予測手法も提案されている。
- ◆ しかし：単一の代表構造に基づく解析であり、RNA 構造のゆらぎ・確率的変動や構造アンサンブルを十分には考慮していない。

### Our idea: Loop profiles

- ◆ 塩基  $i$  が Stem / Hairpin / Internal / Bulge / Multiloop / Exterior (Fig3) のどのループに属するかを表す確率ベクトル = 「ループプロファイル」に注目。
- ◆ エネルギーパラメータに基づいて計算される、構造アンサンブルを反映した指標である。(Fukunaga et al., 2014)

### Aim of this study

- ◆ 大規模 SHAPE データセットを用いてループプロファイルと reactivity の関係を網羅的に解析する。
- ◆ その結果をもとに、①SHAPE-guided 二次構造予測と②エネルギーパラメータ推定への応用を見据えた統計的基盤を築く。

## Materials and methods

### Dataset

RMDB から高信頼度の SHAPE データのみを取得した。(<440nt)

1M7 試薬を用いたデータに限定。1M7 試薬：塩基種類のバイアスが小さい

### ループプロファイルとその計算方法

各塩基  $x_i$  がループ  $loop$  に所在する確率 = 「ループプロファイル  $p(loop|x_i)$ 」を、CapR (Fukunaga et al., 2014) と LinCapR (Otagaki et al., 投稿中) を用いて計算した。(Fig4)

$loop \in \{\text{Stem, Hairpin, Internal, Bulge, Multiloop, Exterior}\}$

Fig5 に示す概略のように、各塩基の reactivity とループプロファイルの値を比較した。 Fig4. CapRによるループプロファイルの例



[https://rmdb.stanford.edu/stanford\\_media/shapereadme\\_rdmdb\\_tg.pdf](https://rmdb.stanford.edu/stanford_media/shapereadme_rdmdb_tg.pdf)

0,1,...  
GGGGGAA A AAACCC C A

7番目の塩基が hairpin loop に所在する確率が 98%

$p(7, \text{hairpin}) = 0.98$

14番目の塩基が stem に所在する確率が 95%

$p(14, \text{stem}) = 0.95$

最小自由エネルギー構造 (最大確率の構造)

Fig4. CapRによるループプロファイルの例

## Results

### Take Home Message: ループプロファイルは塩基対確率単体と比べて、SHAPE reactivity をよりよく反映する構造指標であることを、大規模データで示した。

高 reactivity ほど Stem 確率 ↓, Hairpin 内側確率 ↑  
Exterior / Multiloop は中程度の reactivity の極大

Expectation of  $p(loop|r)$  (ALL, n=4,292,020)

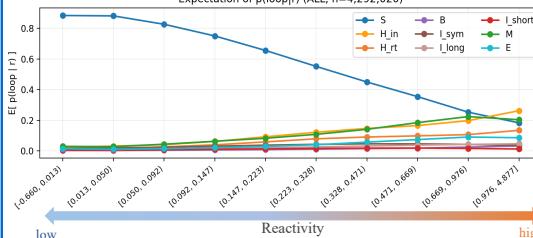


Fig6. SHAPE reactivity を bin に分け、各 bin ごとのループプロファイル平均を示した。Reactivity が高いほど Stem (S) の確率が低下し、hairpin 内側 (H\_in) の確率が上昇する。これらの傾向は右上の先行研究と整合的であり、大規模データでも同様のパターンが確認された。また Exterior loop, Multiloop は中程度の reactivity で最大となることもわかる。

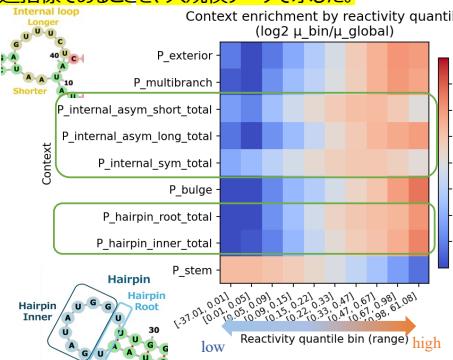


Fig7. 各 reactivity bin におけるループプロファイル平均  $\mu_{loop,r}$  と全塩基での平均  $\mu_{loop}$  の対数オッズ比を示す。Internal asym long (長い方の internal loop), hairpin inner (内側 hairpin), bulge は高 reactivity で、Internal asym short (短い方の internal loop) や対称な internal loop は中程度の reactivity で、それぞれエンリッチしており、ループ種類ごとに異なる reactivity パターンが現れている。

## Discussion

- 比較的長い配列を含む 50,000 以上の配列からなるデータから、網羅的解析によって構造の確率的指標と reactivity の傾向を示した
- ループプロファイルが、単なる塩基対確率よりも多くの reactivity の情報を持つということが示唆された。  
→ 構造プロファイルベースのエネルギーパラメータ推定の基盤として利用できる重要な観察結果である。

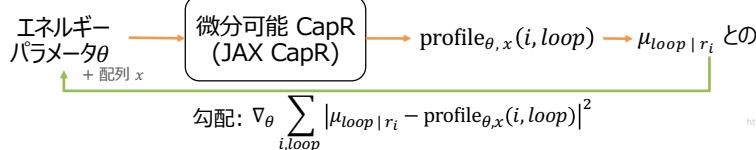


Fig8. ①勾配法と微分可能 CapR によるエネルギーパラメータ推定の概念図

- エネルギーパラメータ推定：各 reactivity の値に対して定まるループプロファイルの平均値  $\mu_{loop|r}$  を正しいループプロファイルだとみなし、それに沿う結果を出力するようなエネルギーパラメータを、微分可能 CapR (JAX-CapR) の微分を計算し勾配法で最適化する。(Fig8)

- SHAPE-guided 二次構造予測：reactivity  $r_i$  が観測された時のループプロファイルの期待値  $\mu_{loop|r_i}$  が高いループ構造  $loop$  をなるべく予測する構造予測。(Fig9)

- 修飾塩基の reactivity の公開データは少ないため、東京大学 定量科学研究所 齊藤研にて実験を実施中。

①,②ともに、本解析によって得られた、reactivity  $r_i$  を観測した時のループ  $loop$  についてのループプロファイルの平均値  $\mu_{loop|r_i}$  を活用する。

$$\sigma^* := \arg \min_{\sigma} (F_x(\sigma) + \lambda E_{s>p(|r|)} [\text{pseudo\_energy}(\sigma, s)]) \\ = \arg \min_{\sigma} \left( F_x(\sigma) + \lambda \sum_{i} \sum_{loop \in L} \mu_{i,loop|r_i} \right)$$

Reactivity  $r_i$  が観測された時のあるループ種類 (例えば hairpin) についてのその塩基のループプロファイルの平均値 (期待値)。  
これが高いほど塩基  $i$  をそのループに割り当てる構造が選ばれやすくなる

Fig9. Reactivity を利用した二次構造予測