

PRESENTATION

PaRNAssus: More insights into RNA probing data

YAO, Hua-Ting

TBI, University of Vienna, Austria

tbi

in collaboration with



Benasque — Jul 29, 2024

- Deigan method, pseudo-energy for stacked position i (Deigan *et al.*, 2009)

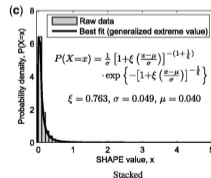
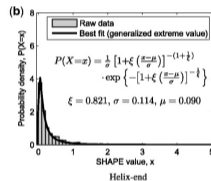
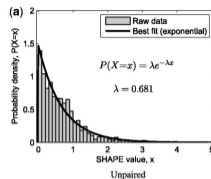
$$\Delta G(i) = m \log(r_i + 1) + b$$

- Probabilistic model (S. Eddy, 2014); RNAprob (F. Deng *et al.*, 2016)

$$\operatorname{argmax}_S \mathbb{P}(S | w) \prod_i \mathbb{P}(r_i | \pi_i)$$

π_i : structural context of position i

- One probing dataset (RNAprob) with reference structures (Deigan *et al.*, 2009; Hajdin *et al.*, 2013; Lavender *et al.*, 2015; F. Deng *et al.*, 2016)



(Z Sükösd *et al.*, 2013)

- Deigan method, pseudo-energy for stacked position i (Deigan *et al.*, 2009)

$$\Delta G(i) = m \log(r_i + 1) + b$$

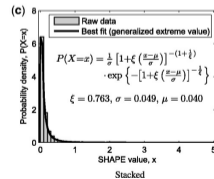
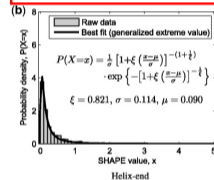
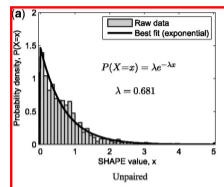
- Probabilistic model (S. Eddy, 2014); RNAprob (F. Deng *et al.*, 2016)

$$\operatorname{argmax}_S \mathbb{P}(S | w) \prod_i \mathbb{P}(r_i | \pi_i)$$

π_i : structural context of position i

- One probing dataset (RNAprob) with reference structures (Deigan *et al.*, 2009; Hajdin *et al.*, 2013; Lavender *et al.*, 2015; F. Deng *et al.*, 2016)

How come that the reactivity distributions of unpaired overlap so much with others?



(Z Sükösd *et al.*, 2013)

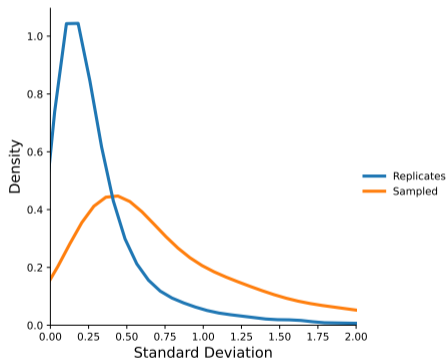


- 26 non-coding RNAs from PDB
 - A: 27.2%, C: 22%, G: 29.5%, U: 21.3%
 - Annotation with FR3D-python (commit: 3c7dc2)
 - Paired: 54.2%, Unpaired: 45.8%
- Probing with 1M7 at 30°C with 5mM Mg²⁺
- **3 replicates**
 - experimental noise
 - systematic differences between positions

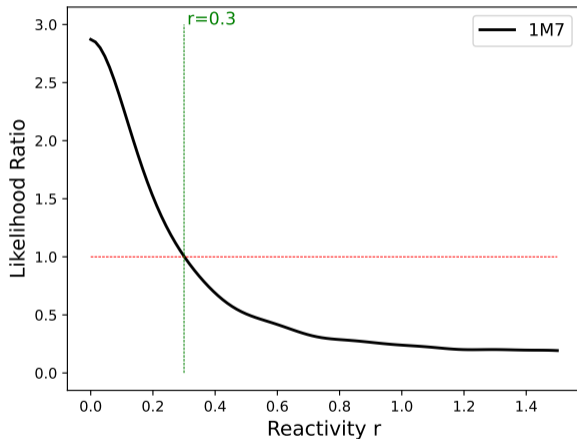


- 26 non-coding RNAs from PDB
 - A: 27.2%, C: 22%, G: 29.5%, U: 21.3%
 - Annotation with FR3D-python (commit: 3c7dc2)
 - Paired: 54.2%, Unpaired: 45.8%
- Probing with 1M7 at 30°C with 5mM Mg²⁺
- **3 replicates**
 - experimental noise
 - systematic differences between positions

standard deviation of 3 reactivity values in unpaired region

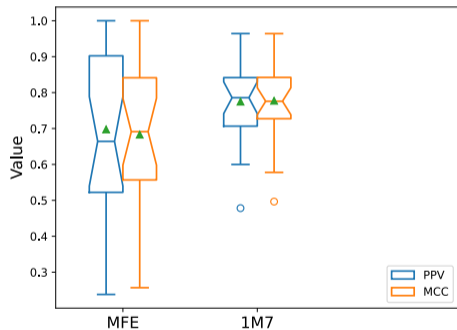


Likelihood ratio $\frac{\mathbb{P}(r|\text{Paired})}{\mathbb{P}(r|\text{Unpaired})}$ (PaRNAssus dataset)



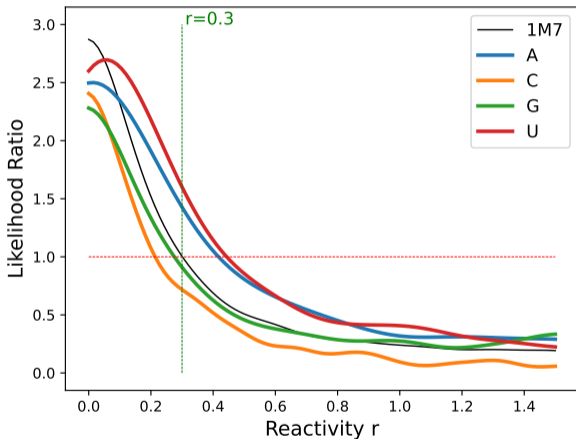
Prediction on RNAProb dataset

$$\operatorname{argmax}_S \mathbb{P}(S | w) \prod_i \mathbb{P}(r_i | \pi_i \in \{\text{Paired, Unpaired}\})$$



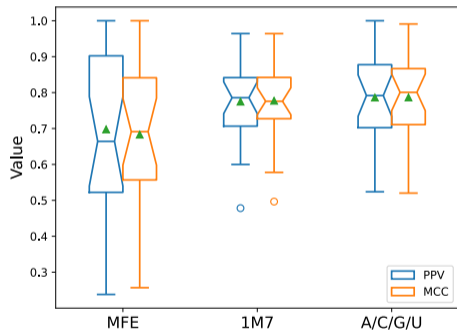
Is reactivity sequence dependent?

Likelihood ratio $\frac{\mathbb{P}(r|\text{Paired})}{\mathbb{P}(r|\text{Unpaired})}$ (PaRNAssus dataset)



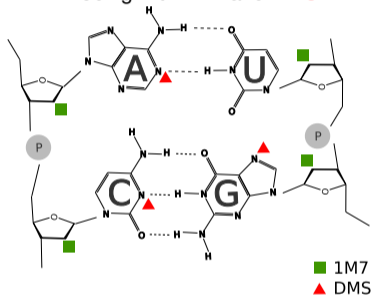
Prediction on RNAprob dataset

$$\operatorname{argmax}_S \mathbb{P}(S | w) \prod_i \mathbb{P}(r_i | \pi_i, w_i \in \{A, C, G, U\})$$



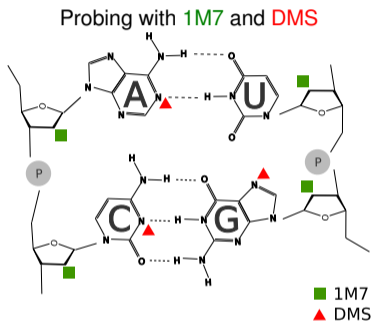
Is reactivity sequence dependent?

Probing with 1M7 and DMS



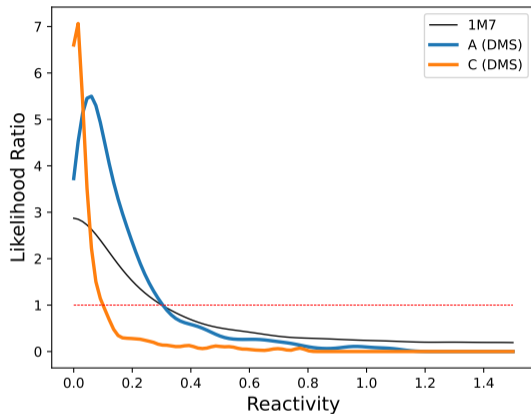
1M7: Measure nucleotide flexibility

DMS: Measure A and C pairedness



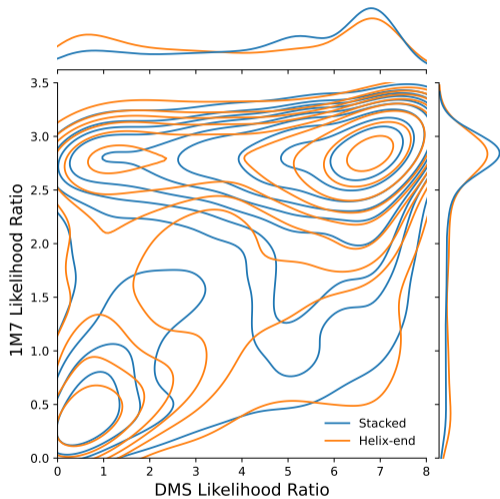
1M7: Measure nucleotide flexibility

DMS: Measure A and C pairedness

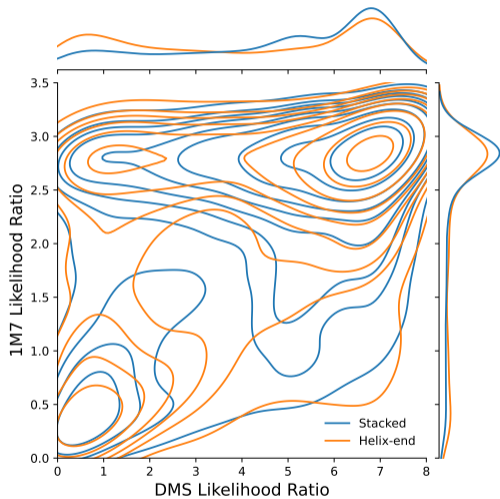


Nucleotide dependency is important for DMS

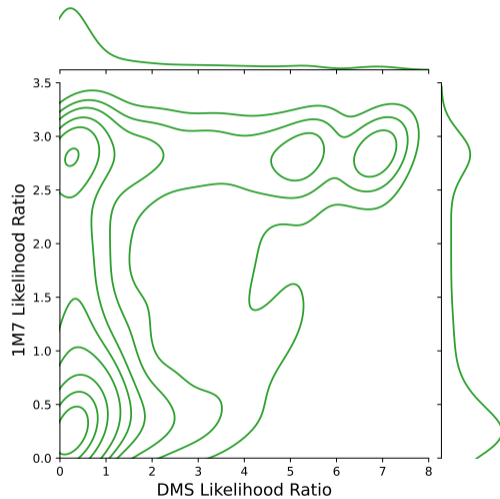
cWW Paired



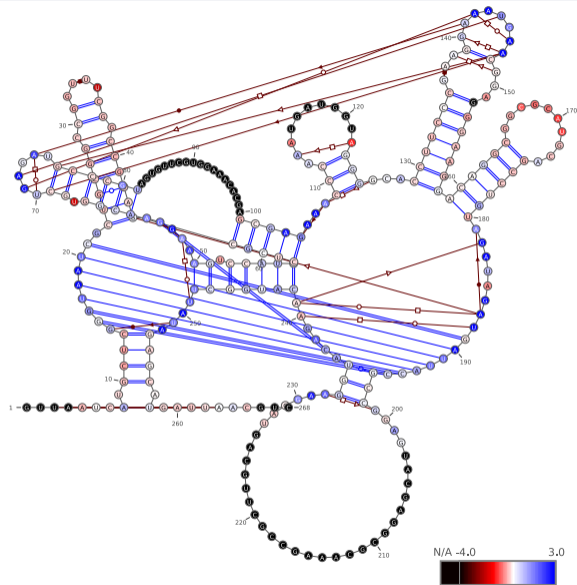
cWW Paired



cWW Unpaired

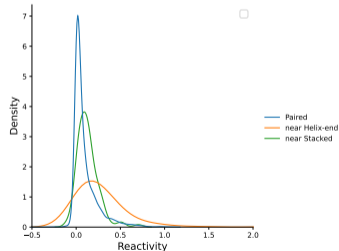


- Hyp: presence of Mg^{2+} forms long range interactions
- 1M7 probing with and without 5mM Mg^{2+}
- Reactivity change $\Delta r_i = r_i(\text{w/o } Mg^{2+}) - r_i(\text{w/ } Mg^{2+})$
→ **Positive change**: position is more reactive w/o Mg^{2+}



RNase P RNA (3DHS chain A)

- PaRNAssus dataset
 - Probing with 1M7 and DMS
 - with and without Mg^{2+}
 - different temperatures $30^{\circ}C$ and $55^{\circ}C$
- Probing data contains more information than paired/unpaired
 - Sequence dependence crucial for DMS
 - Probing under different conditions reveals long range interactions and PKs
 - and other structural information, *e.g.* near interactions

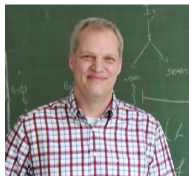


- Integration into structure prediction tools, *e.g.* RNAPKplex for PK prediction

tbi



CiTCoM



FWF Autonome Nationalde de la Recherche ANR