

39TH TBI WINTERSEMINAR

Something about probing

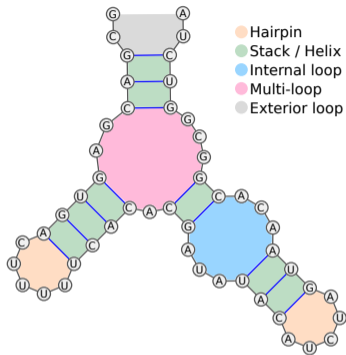
Analysis report of PaRNAssus reactivity data

YAO, Hua-Ting

TBI, University of Vienna, Austria

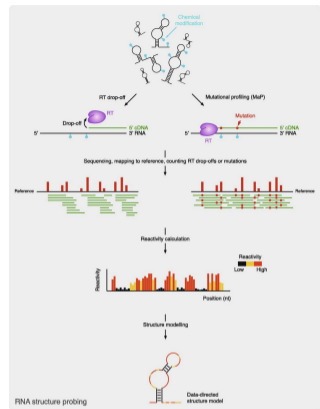
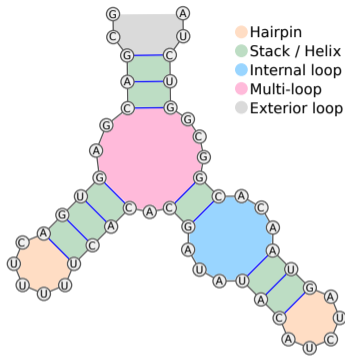
The logo for TBI (Theoretical Bioinformatics Institute) is written in a stylized, lowercase, cursive font.

Bled — Feb 13, 2024



- The free energy $\mathcal{E}(w, S)$ of secondary structure S given sequence w is

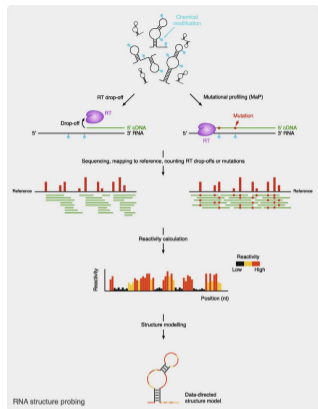
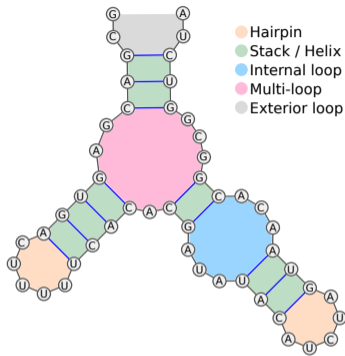
$$\mathcal{E}(w, S) = \Delta \left(\text{hairpin} \right) + \Delta \left(\text{stack} \right) + \Delta \left(\text{internal loop} \right) + \Delta \left(\text{multi-loop} \right) + \dots$$



(S Aviran & D Incarnato, 2022)

- The free energy $\mathcal{E}(w, S)$ of secondary structure S given sequence w is

$$\mathcal{E}(w, S) = \Delta \left(\text{Hairpin} \right) + \Delta \left(\text{Stack / Helix} \right) + \Delta \left(\text{Internal loop} \right) + \Delta \left(\text{Multi-loop} \right) + \dots$$



(S Aviran & D Incarnato, 2022)

- The free energy $\mathcal{E}(w, S)$ of secondary structure S given sequence w is

$$\mathcal{E}(w, S) = \Delta \left(\text{hairpin} \right) + \Delta \left(\text{stack/helix} \right) + \Delta \left(\text{internal loop} \right) + \Delta \left(\text{multi-loop} \right) + \dots$$

- Reactivity r_i : value in $[0, > 1]$ at position i , low: paired, high: unpaired

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

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

- Probabilistic model (S Eddy, 2014; 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

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

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

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

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

π_i : structural context of position i

- **Nucleotide**-dependence?
- Structural dependence?
 - Paired: stacked, helix-end . . .
 - Unpaired: loop type, non-canonical basepairs . . .

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

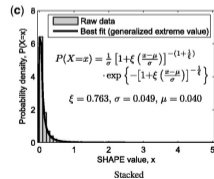
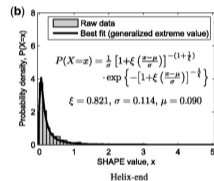
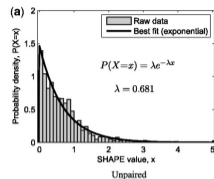
$$\Delta G(i) = m \log(r_i + 1) + b$$

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

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

π_i : structural context of position i

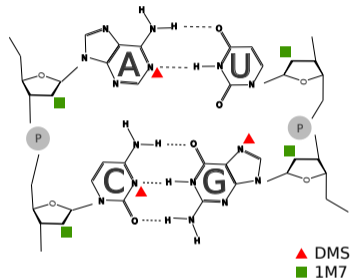
- Nucleotide**-dependence?
- Structural dependence?
 - Paired: stacked, helix-end ...
 - Unpaired: loop type, non-canonical basepairs ...



(Z Sükösd *et al.*, 2013)

- Joint project with B. Sargueil and Y. Ponty (France)
- 36 non-coding RNAs from PDB
 - A: 26%, C: 23%, G: 29%, U: 22%
 - Paired: 55%, Unpaired: 45%

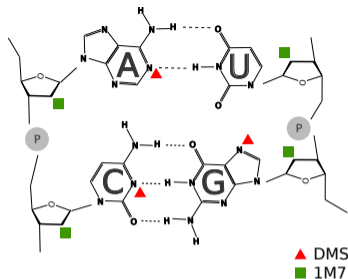
- Joint project with B. Sargueil and Y. Ponty (France)
- 36 non-coding RNAs from PDB
 - A: 26%, C: 23%, G: 29%, U: 22%
 - Paired: 55%, Unpaired: 45%
- Probing with **DMS** and **1M7**
- 30° C with Mg²⁺
- 3 replicates



DMS: Measure A and C pairedness

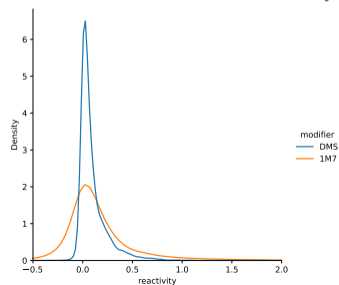
1M7: Measure nucleotide flexibility

- Joint project with B. Sargueil and Y. Ponty (France)
- 36 non-coding RNAs from PDB
 - A: 26%, C: 23%, G: 29%, U: 22%
 - Paired: 55%, Unpaired: 45%
- Probing with **DMS** and **1M7**
- 30°C with Mg²⁺
- 3 replicates

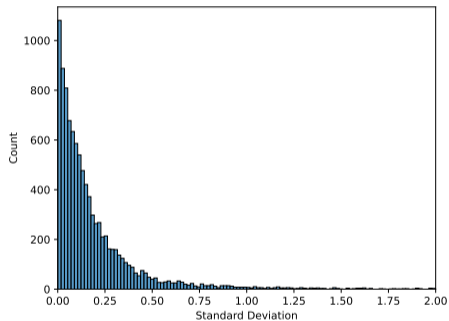


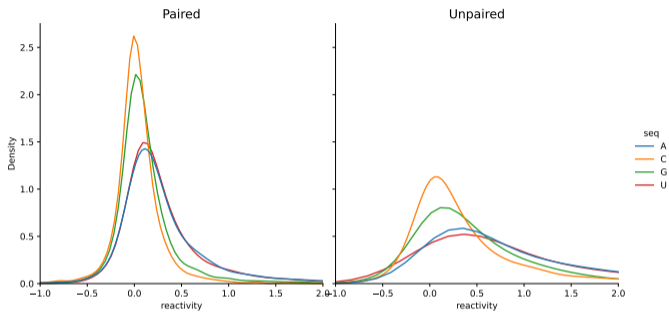
DMS: Measure A and C pairedness

1M7: Measure nucleotide flexibility

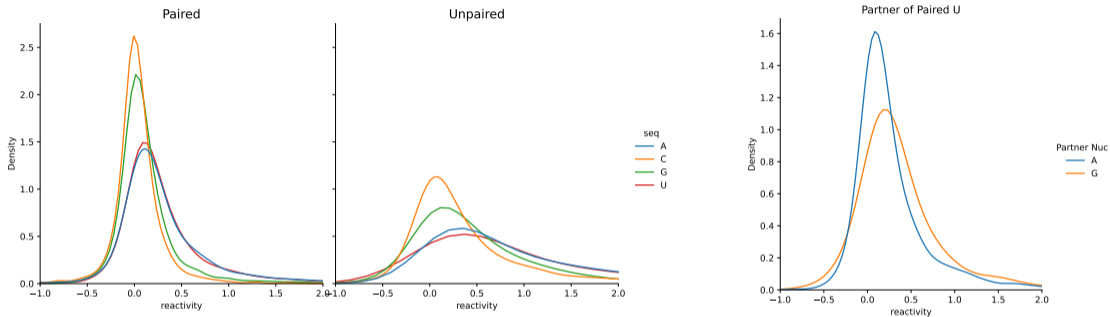


- Joint project with B. Sargueil and Y. Ponty (France)
- 36 non-coding RNAs from PDB
 - A: 26%, C: 23%, G: 29%, U: 22%
 - Paired: 55%, Unpaired: 45%
- Probing with **DMS** and **1M7**
- 30°C with Mg^{2+}
- 3 replicates





- A and U are more flexible



- A and U are more flexible
- AU basepair is less flexible

- RNAProb dataset (Deng *et. al.*, 2016): 19 RNAs (< 600 nts) + 4 rRNAs (> 1 000 nts)
- Probabilistic model (Prob)

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

- RNAProb dataset (Deng *et. al.*, 2016): 19 RNAs (< 600 nts) + 4 rRNAs (> 1 000 nts)
- Probabilistic model (Prob), **AU/CG**

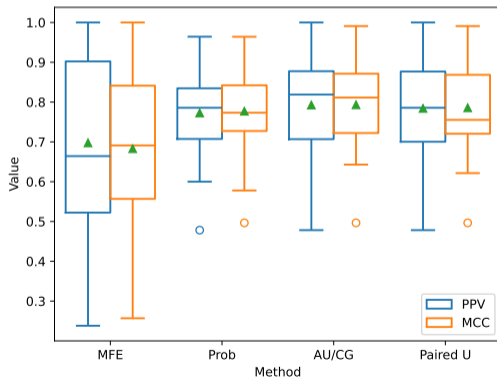
$$\operatorname{argmax}_S \mathbb{P}(S \mid w) \prod_i \mathbb{P}(r_i \mid \pi_i, w_i \in \{\mathbf{AU}, \mathbf{CG}\})$$

- RNAProb dataset (Deng *et. al.*, 2016): 19 RNAs (< 600 nts) + 4 rRNAs (> 1 000 nts)
- Probabilistic model (Prob), AU/CG, Paired U

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

- RNAprob dataset (Deng *et. al.*, 2016): 19 RNAs (< 600 nts) + 4 rRNAs (> 1 000 nts)
- Probabilistic model (Prob), **AU/CG**, **Paired U**

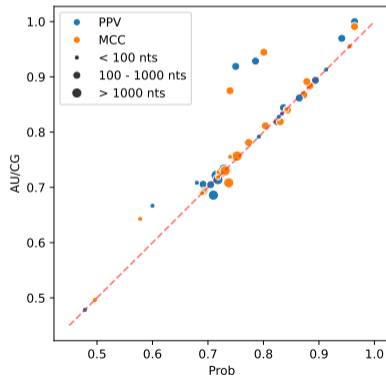
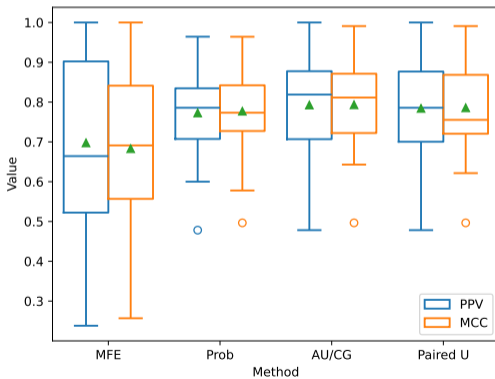
$$\operatorname{argmax}_S \mathbb{P}(S | w) \mathbb{P}(r_i | \pi_i = \text{Paired}, w_i = \text{U}, w_j \in \{\text{A}, \text{G}\}) \prod_i \mathbb{P}(r_i | \pi_i)$$

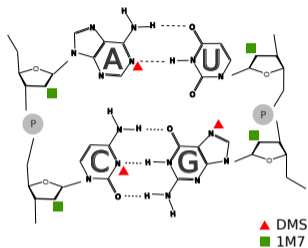


Structure Prediction on RNAprob Dataset

- RNAprob dataset (Deng *et. al.*, 2016): 19 RNAs (< 600 nts) + 4 rRNAs (> 1 000 nts)
- Probabilistic model (Prob), **AU/CG**, **Paired U**

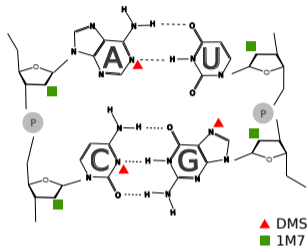
$$\operatorname{argmax}_S \mathbb{P}(S | w) \mathbb{P}(r_i | \pi_i = \text{Paired}, w_i = \text{U}, w_j \in \{\text{A}, \text{G}\}) \prod_i \mathbb{P}(r_i | \pi_i)$$





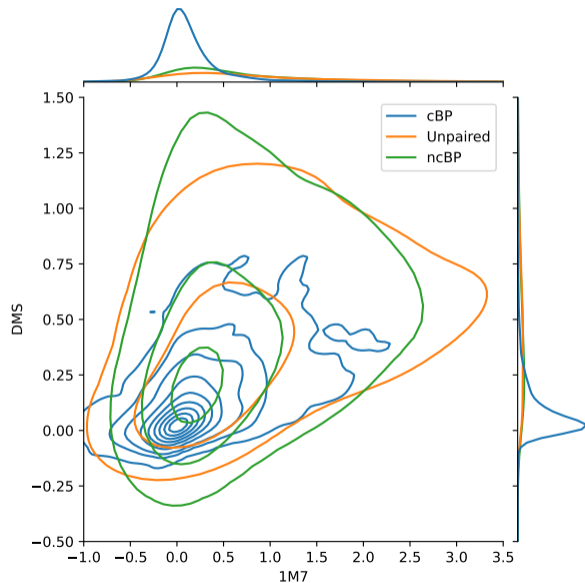
DMS: Measure A and C pairedness

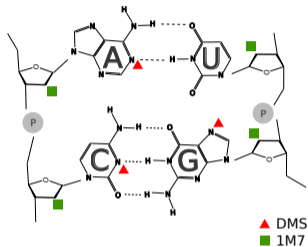
1M7: Measure nucleotide flexibility



DMS: Measure A and C pairedness
1M7: Measure nucleotide flexibility

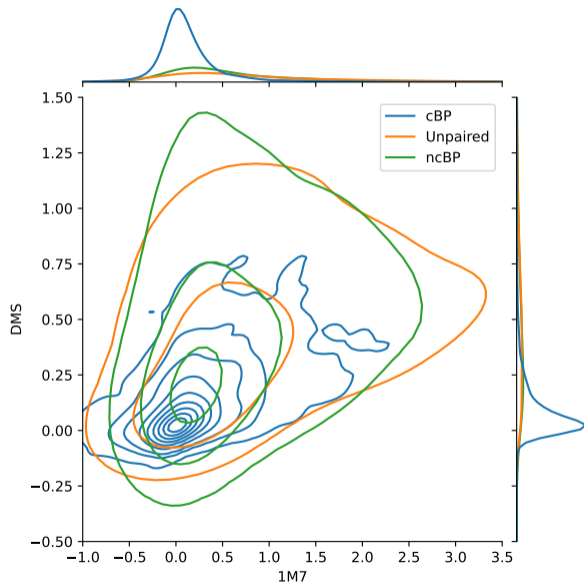
low / low → paired
high / high → unpaired



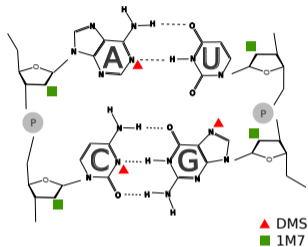


DMS: Measure A and C pairedness
1M7: Measure nucleotide flexibility

low / low → paired
high / high → unpaired
high / low → constrained

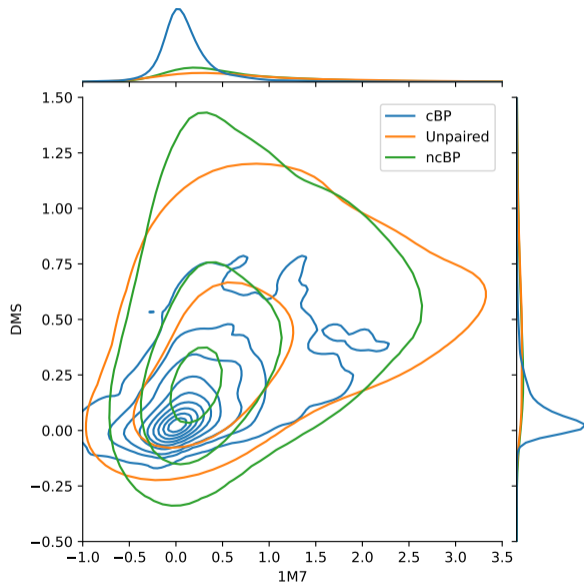


Incorporating DMS and 1M7 reactivity data

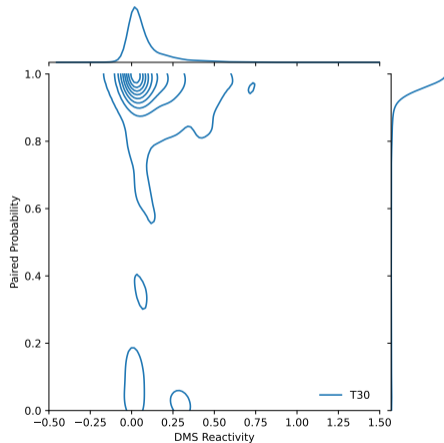


DMS: Measure A and C pairedness
1M7: Measure nucleotide flexibility

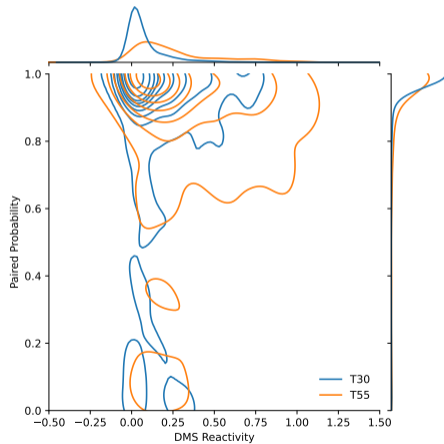
- low / low → paired
- high / high → unpaired
- high / low → constrained
- low / high → ???



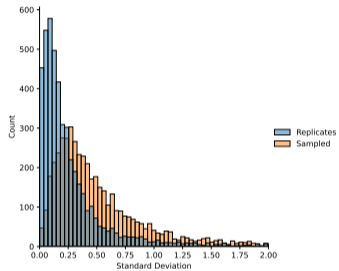
- Temperature increases from 30°C to 55°C
- Compare paired probability and reactivity of each position
→ reference structure as hard constraint



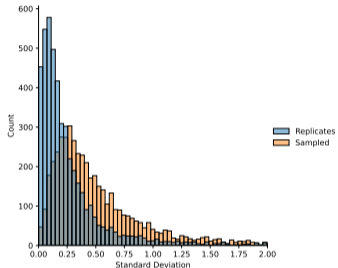
- Temperature increases from 30°C to 55°C
- Compare paired probability and reactivity of each position
→ reference structure as hard constraint



- Structural dependence on unpaired region, such as loop type, 3D motif etc.



- Structural dependence on unpaired region, such as loop type, 3D motif etc.



- Compare with reactivity data w/o Mg^{2+} to identify pseudoknot
- Machine learning to explore hidden features

tbi



CiTCoM

