Computational methods for comparing and integrating multiple probing assays to predict RNA secondary structure

Afaf Saaidi1,2,3, Delphine Allouche3,4
Supervisors: Yann Ponty1,2,3, Bruno Sargueil3,4, Mireille Regnier1,2
1-AMIB team, Inria Saclay 2-LIX 3-CNRS 4-Laboratoire de cristallographie et RMN Biologiques, Paris Descartes

1-Introduction
- RNA is key to understand many biological processes.
- RNA maintains a stable tertiary structure.
- The determination of the structure allows understanding its operating mechanism.
- We study the 444nt long VIH1 Gag-IRES.

RNA Structure determination
- 3D structure can be resolved experimentally [remains expensive and time-consuming].
- Computational methods allow to have accurate secondary structure predictions (PPV ≈ 75%). Less accurate predictions for long RNA.
+ Experimental Data [Chemical SHAPE \ Enzymatic] improve predictions.

State of the art
Evolution of computational approaches to predict the RNA secondary structure:

Objectives
- Apply sampling approach with SHAPE data to predict the RNA secondary structure.

2-Material & Methods
2-1 Experimental data
SHAPE-Map experiments
High Throughput Sequencing

SHAPE reactivity calculation

\[ \text{Reactivity}(n) = \frac{\text{max}_{\text{SHAPE}}(n) - \text{max}_{\text{Control}}(n)}{\text{max}_{\text{Denatured}}(n)} \] [3]

Boltzmann probability to observe a structure \( S \):

\[ P(S) = \frac{1}{Z} e^{-\frac{U(S)}{kT}} \]

\( Z \) the partition function: \( Z = \sum e^{-\frac{U(S)}{kT}} \).

2-2 Sampling/Clustering workflow

experimental data from different conditions

Data processing

Structure sampling

Set of ensemble structures

Clustering[Affinity propagation]

Optimal clusters?

Coherence Diversity Stability

Maximization => Pareto Frontier

Optimal Centroid Structures

3-Results
Optimal centroid structures from 140 [8000 structures]

4-Conclusion & perspectives
- We have obtained a set of models supported by our integrative approach, those models are subject to validation.
- Some of centroid structures have shown high compatibility with existing proposed structural models.
- We will extend the approach to the simultaneous analysis of probing data for a set of RNA variants.

References

Acknowledgments
PhD funded by the "Fondation pour la Recherche Medicale".

Contact address
saaidi.afaf@polytechnique.edu