



HAL
open science

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

Afaf Saaidi, Delphine M Allouche, Yann Ponty, Bruno M Sargueil, Mireille M Regnier

► **To cite this version:**

Afaf Saaidi, Delphine M Allouche, Yann Ponty, Bruno M Sargueil, Mireille M Regnier. Computational methods for comparing and integrating multiple probing assays to predict RNA secondary structure . Doctorial journey Interface, Ecole polytechnique, Palaiseau, Nov 2016, Palaiseau, France. 10. hal-01558227

HAL Id: hal-01558227

<https://hal.science/hal-01558227>

Submitted on 7 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Afaf Saaidi^{1,2,3}, Delphine Allouche^{3,4}

Supervisors: Yann Ponty^{1,2,3}, Bruno Sargueil^{3,4}, Mireille Regnier^{1,2}

1-AMIB team, Inria Saclay 2-LIX, 3-CNRS, 4-Laboratoire de cristallographie et RMN Biologiques, Paris Descartes

1-Introduction

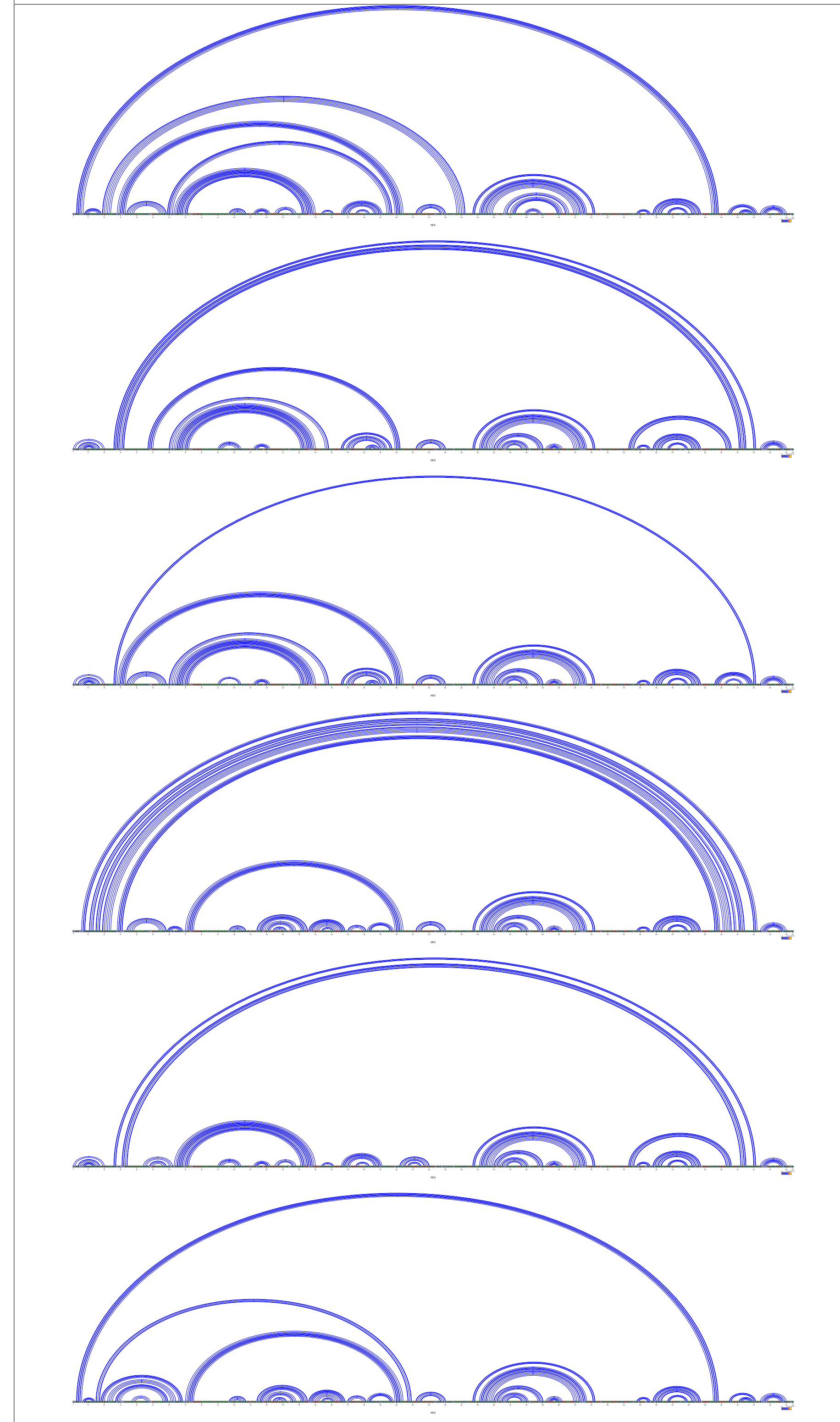
- RNA is key to understand many biological processes.
- RNA maintains a stable tertiary structure.
- The determination of the structure allows understanding its operating mechanism.
- We study the 444nt long **VIH1 Gag-IRES**.

RNA Structure determination

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

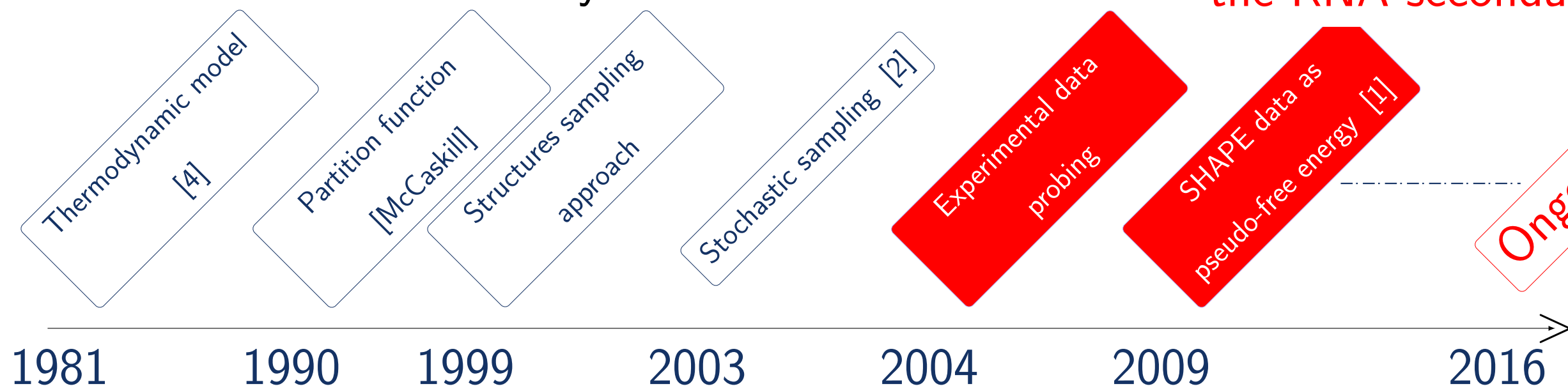
3-Results

Optimal centroid structures from 140. [8000 structures]



State of the art

Evolution of computational approaches to predict the RNA secondary structure:



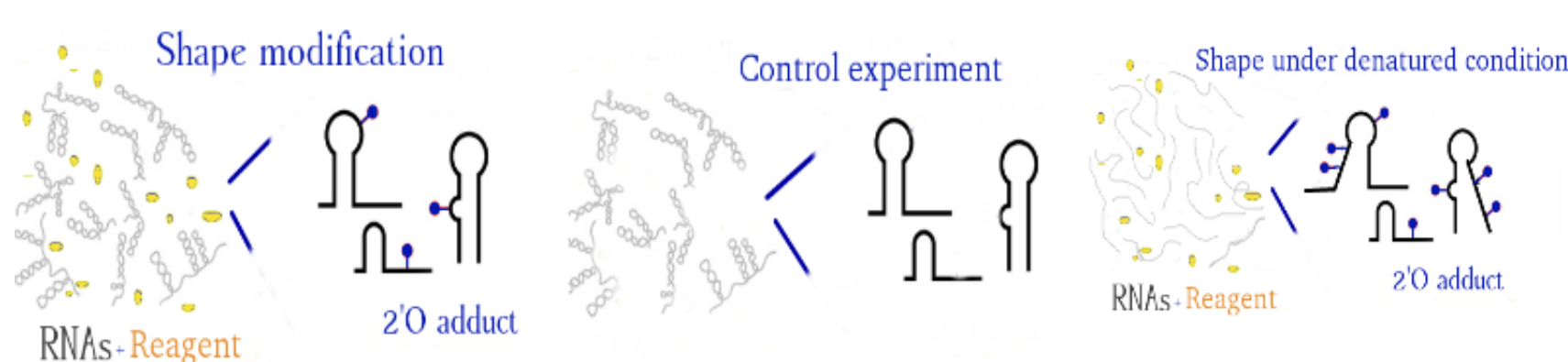
Objectives

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

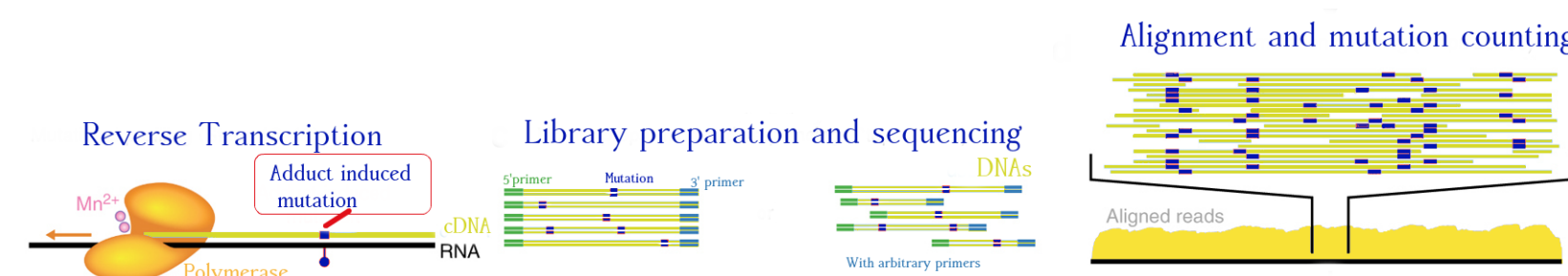
2-Material & Methods

2-1 Experimental data

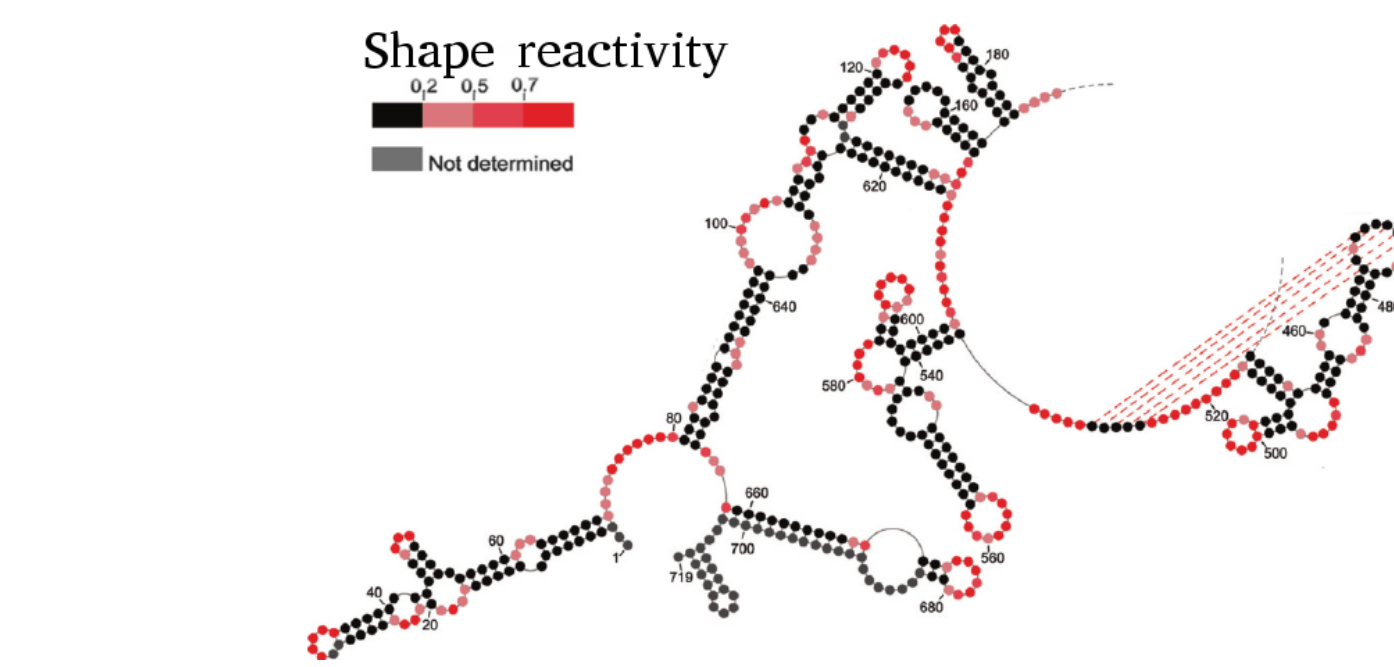
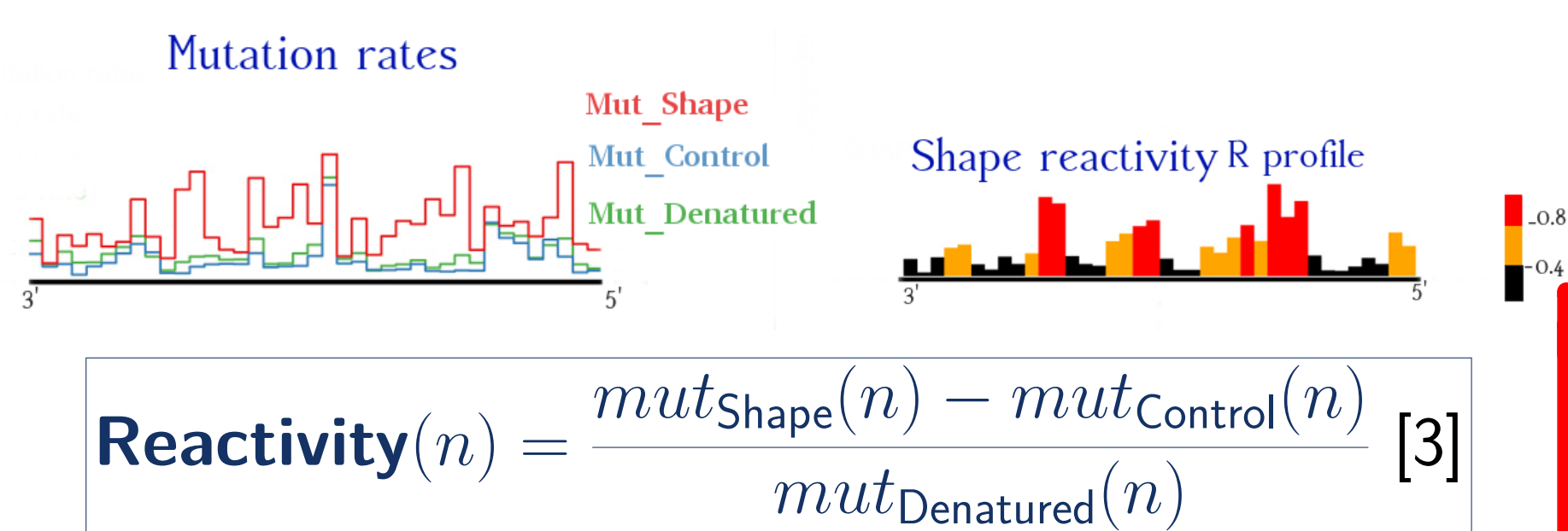
SHAPE-Map experiments



High Throughput Sequencing



SHAPE reactivity calculation



V-Enzymatic cleavage targets paired nucleotides.
T-Enzymatic cleavage reveals unpaired nucleotides.

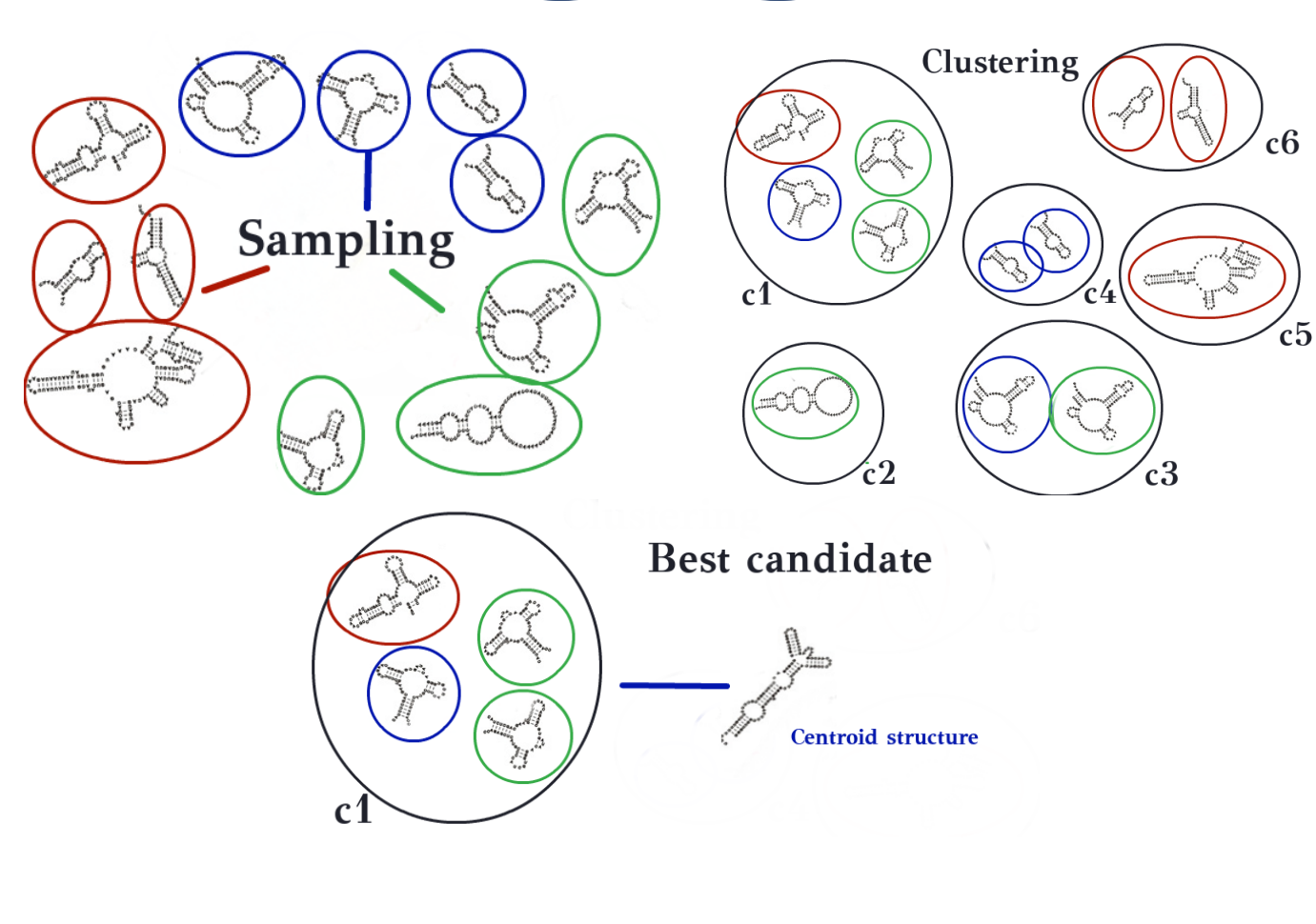
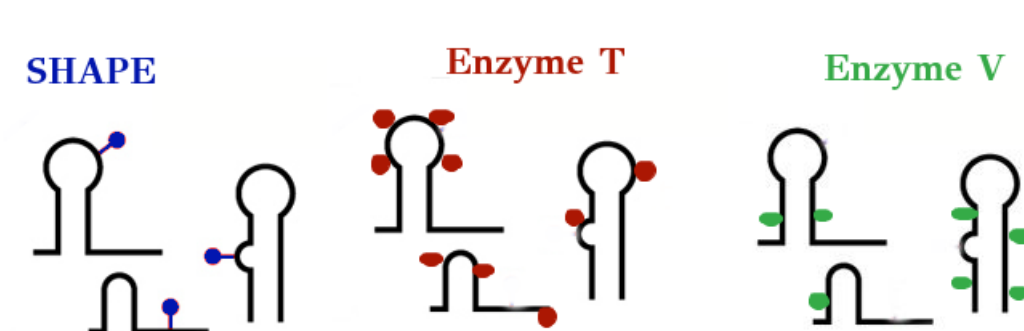
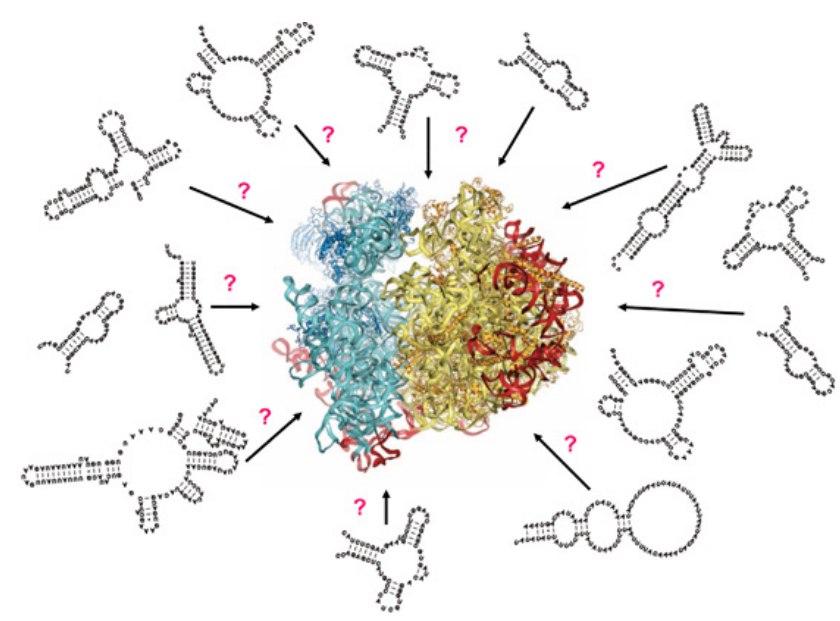
4-Conclusion & perspectives

- 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.

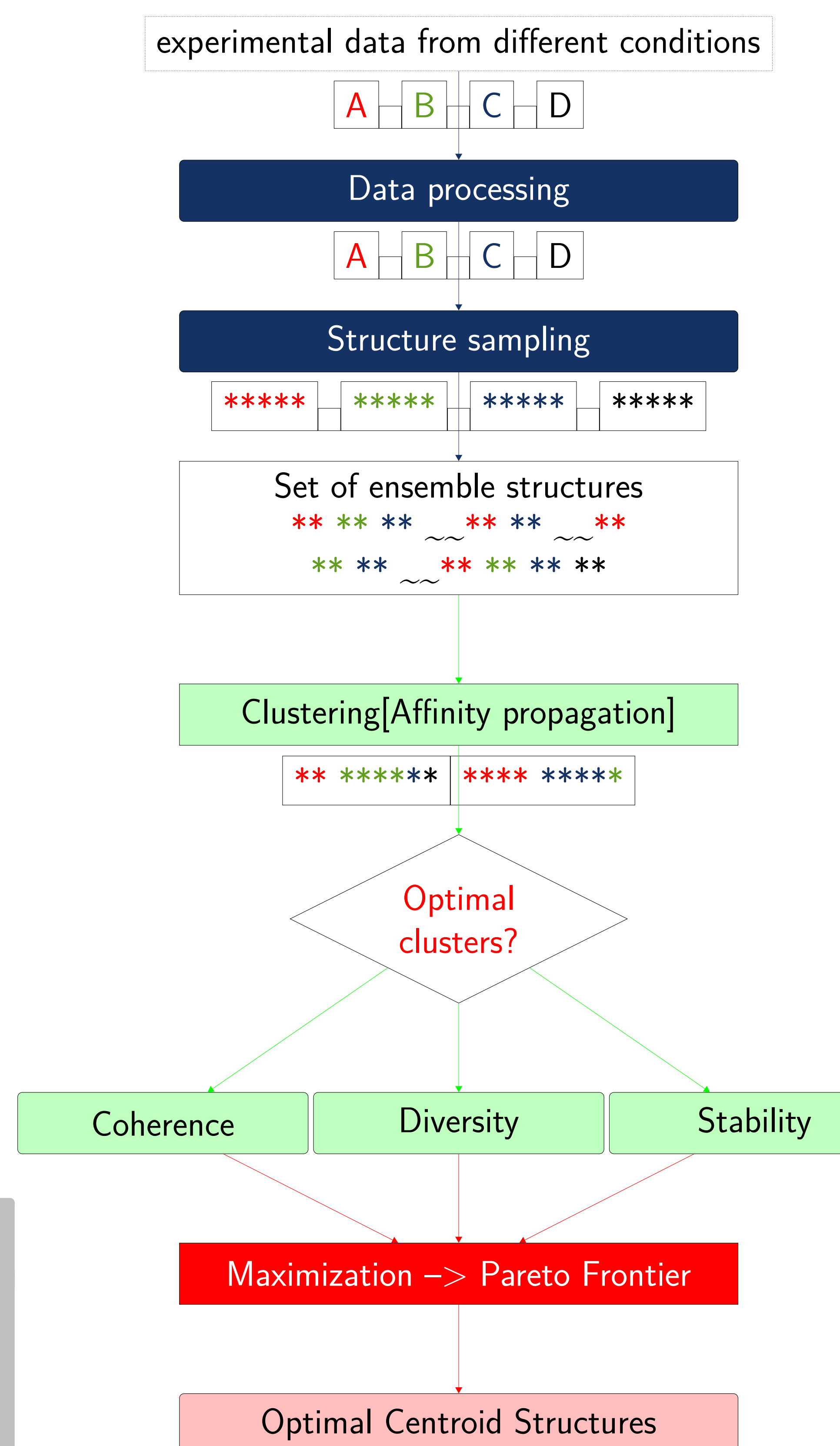
Boltzmann probability to observe a structure S :

$$P(S) = \frac{e^{-\frac{E_S}{kT}}}{Z}$$

with Z the partition function: $Z = \sum_{S \in \mathcal{S}} e^{-\frac{E_S}{kT}}$



2-2 Sampling/Clustering workflow



References

- [1] Katherine E. Deigan, Tian W. Li, David H. Mathews, and Kevin M. Weeks. Accurate SHAPE-directed RNA structure determination. *Nucleic Acids Research*, 37(1):97–102, jan 2009.
- [2] Lawrence CE. Ding. A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Research*, 31(24):7280–7301, 2003.
- [3] Matthew J Smola, Gregory M Rice, Steven Busan, Nathan A Siegfried, and Kevin M Weeks. Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nature protocols*, 10(11):1643–1669, 2015.
- [4] Michael Zuker and Patrick Stiegler. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9(1):133–148, jan 1981.

Acknowledgments

PhD funded by the "Fondation pour la Recherche Médicale".

Contact address

saaidi.afaf@polytechnique.edu

For a cluster \mathcal{C}

- Coherence** \equiv Base pairs **Mean Distance** in \mathcal{C} .
- Diversity** \equiv Number of **present conditions**.
- Stability** \equiv Sum of **Boltzmann probabilities**.