Bruno Sargueil

bruno.sargueil@u-paris.fr

Université Paris Cité

UMR 8038 - CiTCoM
Structure and translation of viral RNA

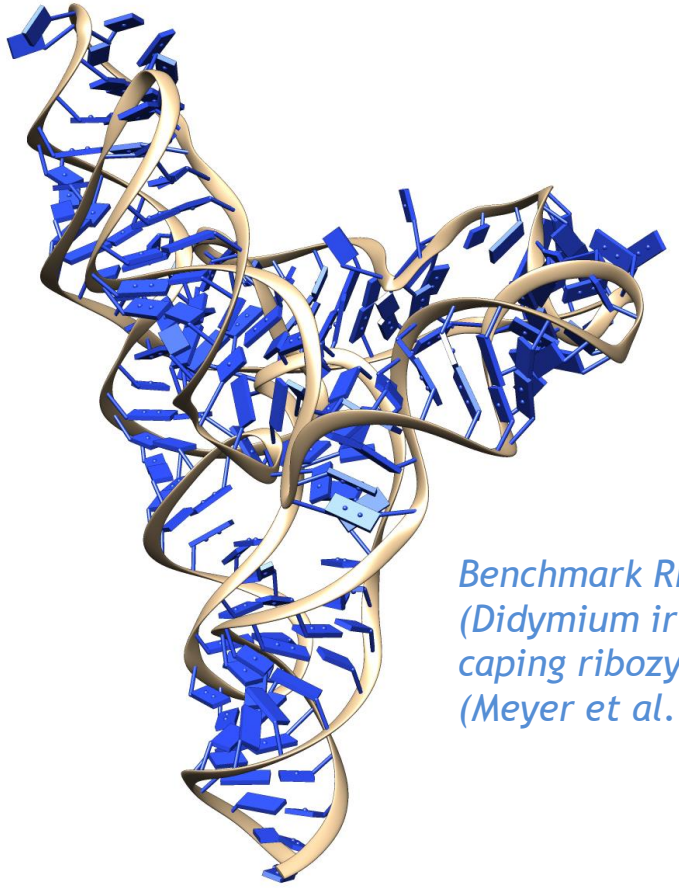
Studying RNA structure dynamics and interactions with SHAPE

"Structure and topology of RNA in living systems"
Trento, 1st of Feruary 2023

Modeling RNA 3D structure

5′GGUUGGGUUGGGAAGUAUCAUGGCUAAUCACC
AUGAUGCAAUCGGGUUGAACACUUAAUUGGGUUA
AAACGGUGGGGGACGAUCCCGUAACAUCCGUCCU
AACGGCGACAGACUGCACGGCCCUGCCUCUUAGG
UGUGUUCAAUGAACAGUCGUUCCGAAAGGAAGCA
UCCGGUAUCCCAAGACAAUC 3′

sequence



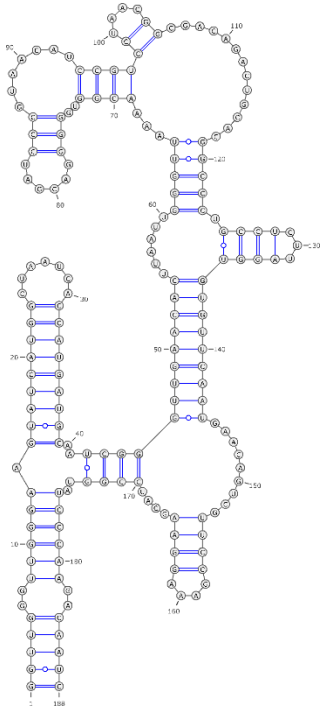
Benchmark RNA : Di-LcRz (Didymium iridis lariat caping ribozyme) (Meyer et al. 2014)

3D structure

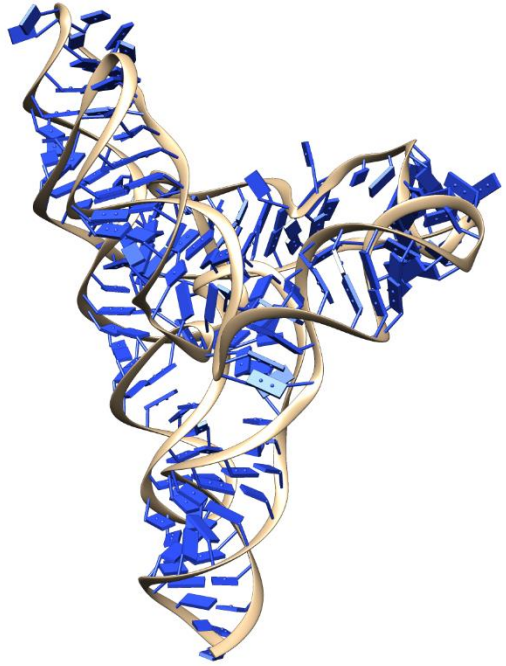
cnrs Université Paris Cité

RNA 3D modeling process

5′GGUUGGGUUGGGAAGUAUCAUGGC
UAAUCACCAUGAUGCAAUCGGGUUGA
ACACUUAAUUGGGUUAAAACGGUGGG
GGACGAUCCCGUAACAUCCGUCCUAA
CGGCGACAGACUGCACGGCCCUGCCU
CUUAGGUGUGUUCAAUGAACAGUCGU
UCCGAAAGGAAGCAUCCGGUAUCCCA
AGACAAUC 3′

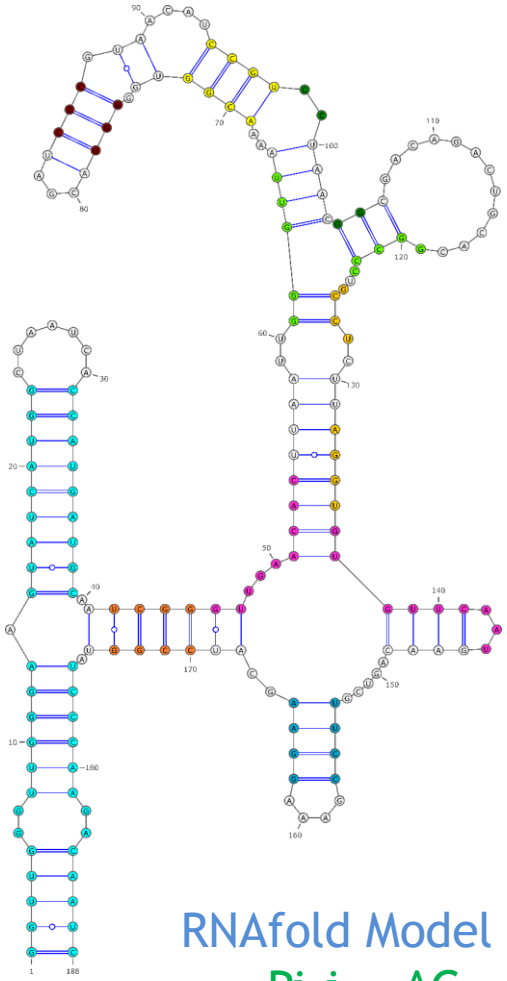


Secondary structure



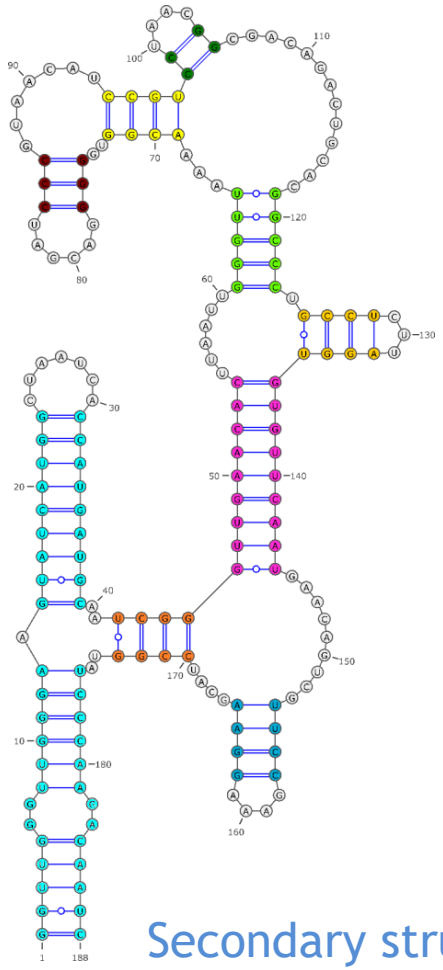
Secondary structure modeling

Most successful algorithms are based on a thermodynamic and probabilistic approach (RNAstructure, RNAFold)



RNAfold Model

Pi,j ΔG_0



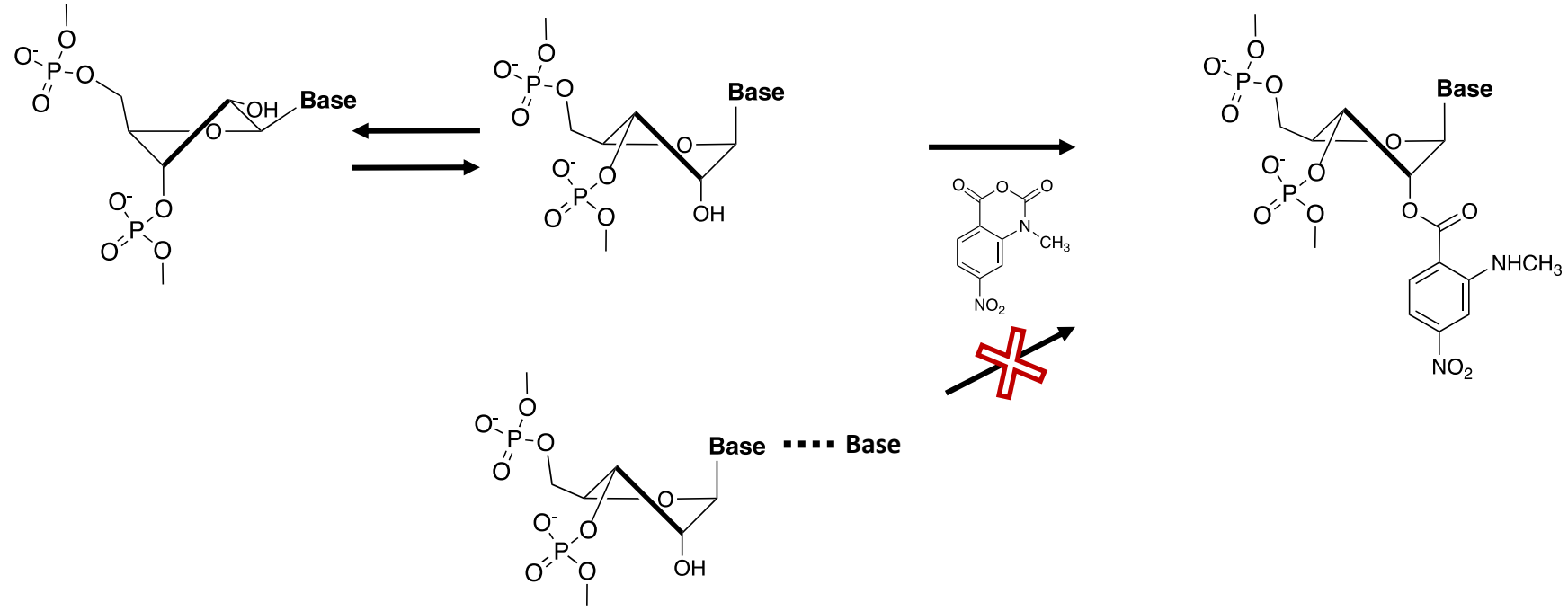
Secondary structure

Most models obtained for RNA over 100-200 nt are not accurate

cnrs Université Paris Cité

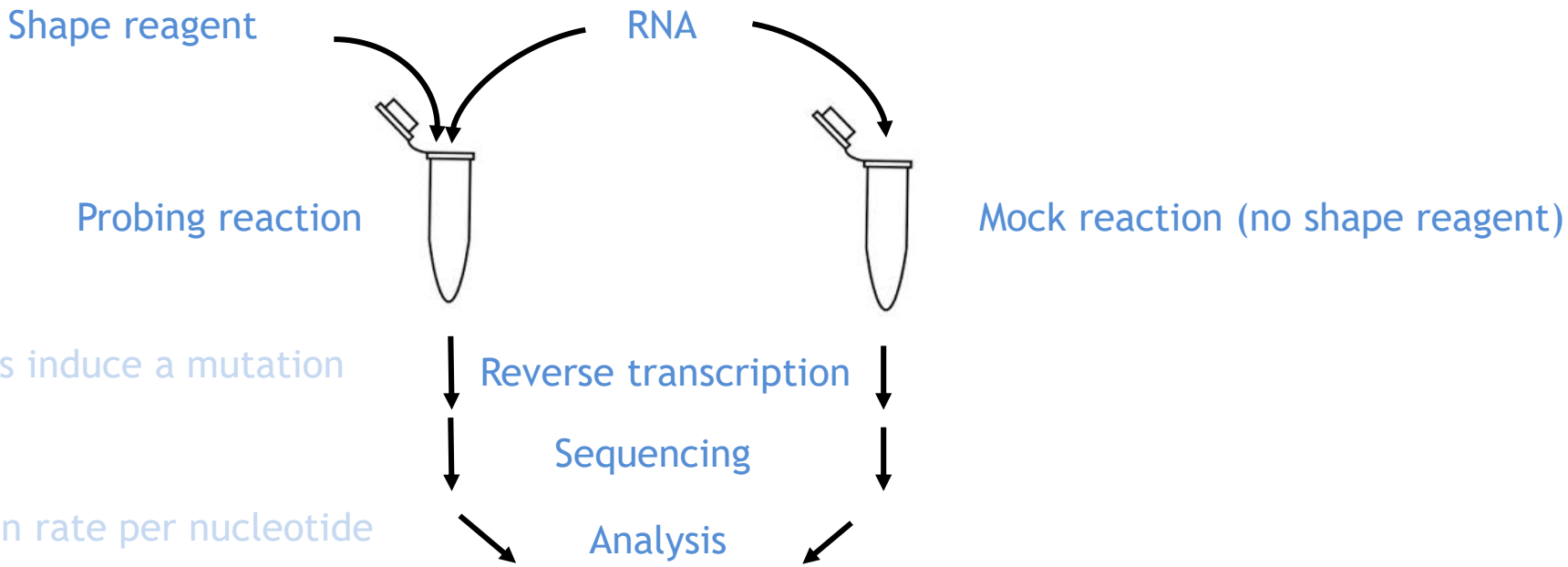
RNA structure chemical probing - SHAPE

The ribose of single strand nucleotides is in a dynamic equilibrium



1. **SHAPE probe senses** ribose flexibility which in turn reflects the **base mobility**
2. Only **single strand nucleotides are reactive**

Structure probing experiment (SHAPE-Map)

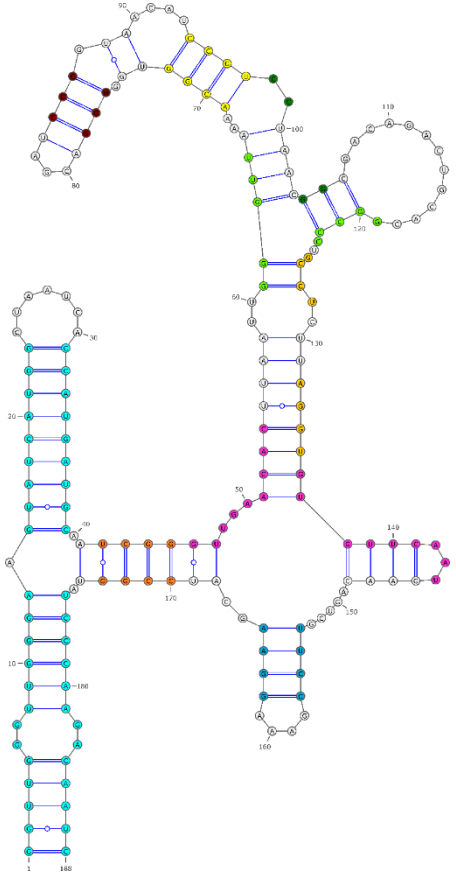


Reactivity Map

May be used to bias the secondary structure prediction

cnrs Université Paris Cité

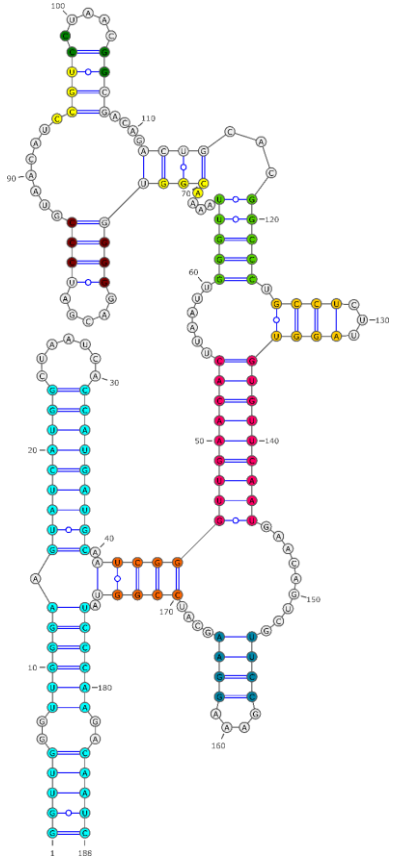
Integrating probing data



RNAfold

Pi,j

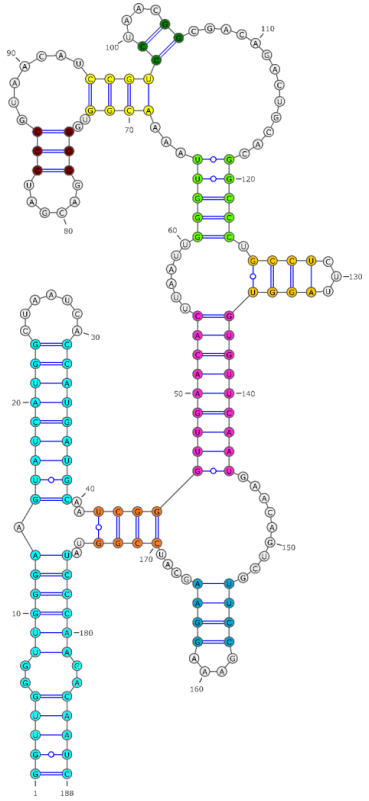
ΔG_0



RNAfold with probing constraints

Pi,j

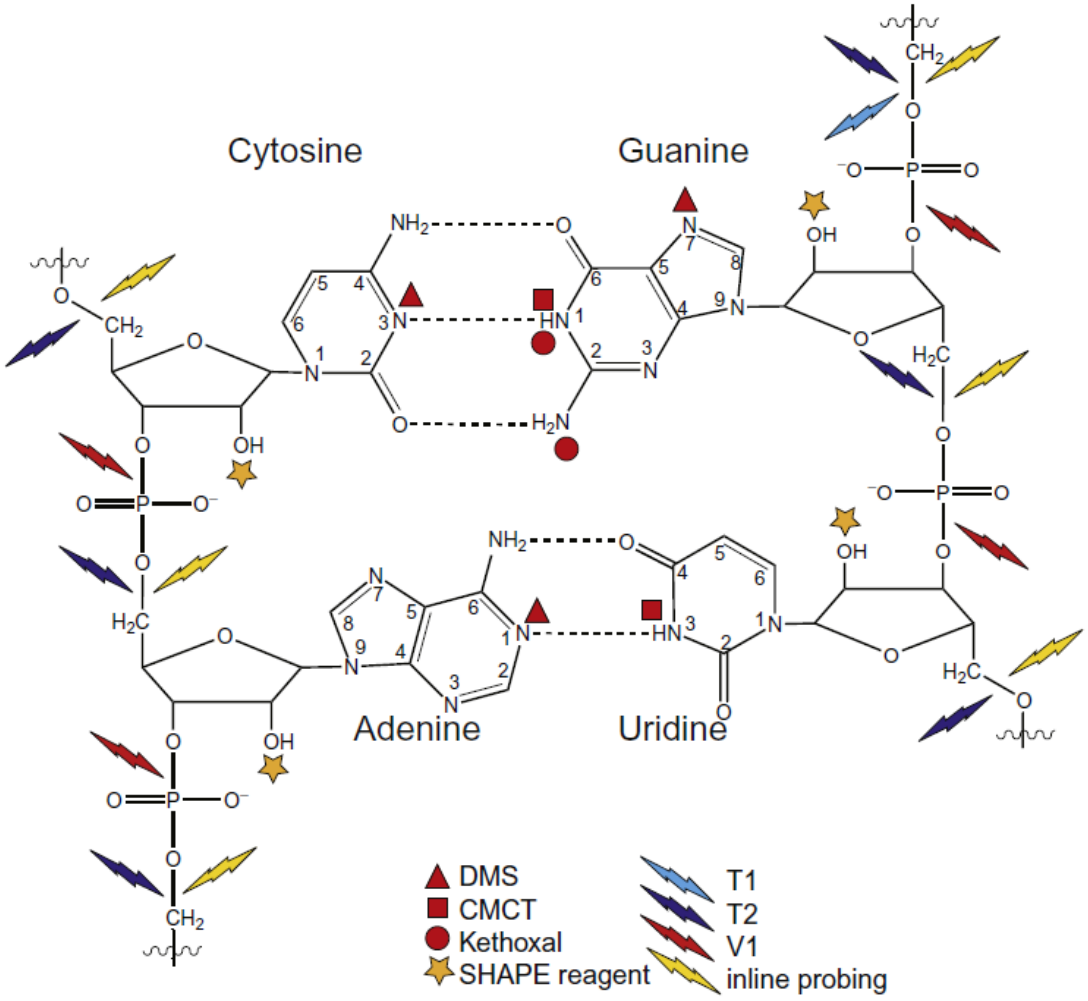
Pseudo $\Delta G'_0$ $\Delta G'_0 = f(\Delta G_0$, réactivity)



2D structure

cnrs Université Paris Cité

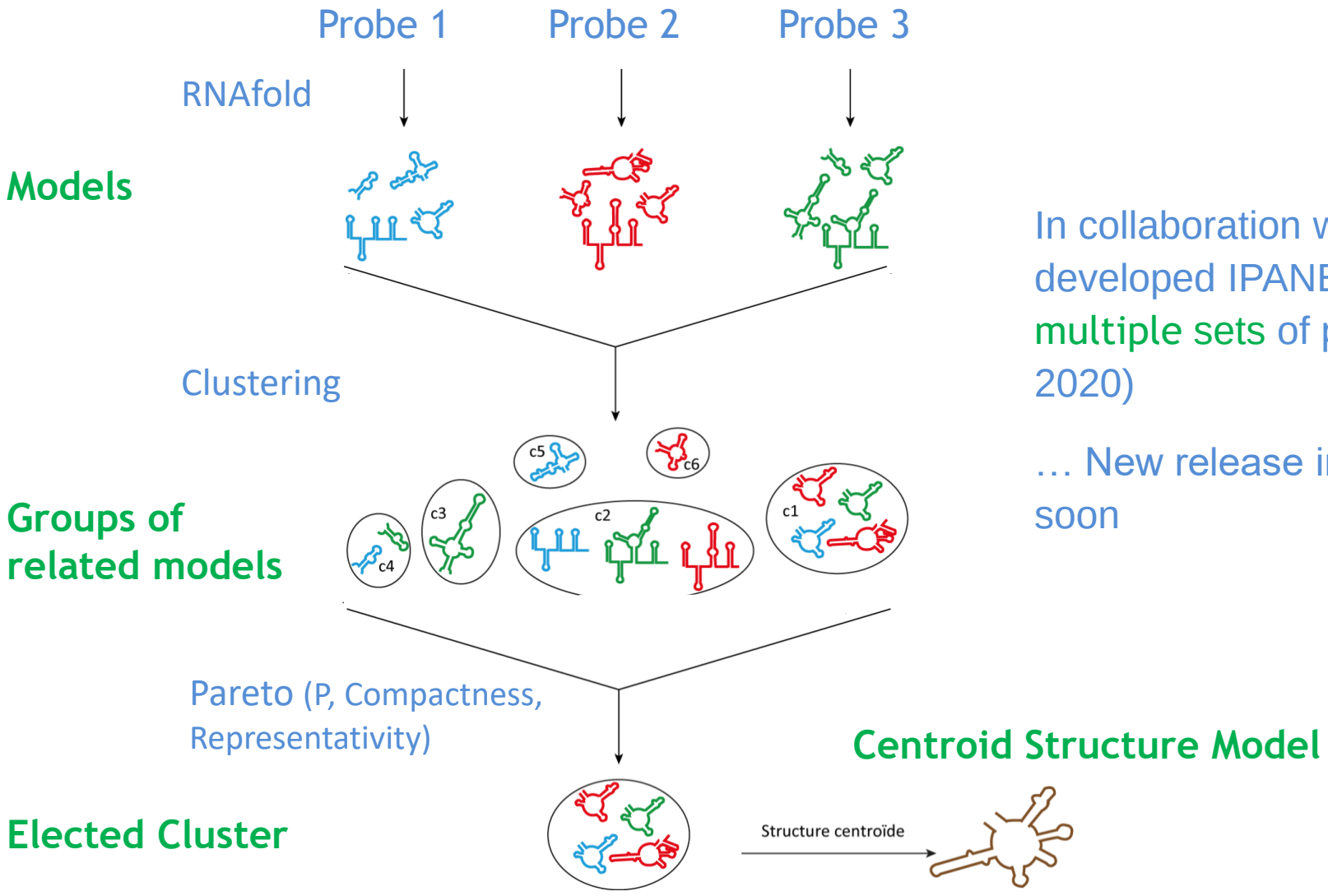
Many probes have been developped



Each probe reports **specific information** but modeling software can take into account only **one data set**

cnrs Université Paris Cité

Integrating multiple probing data

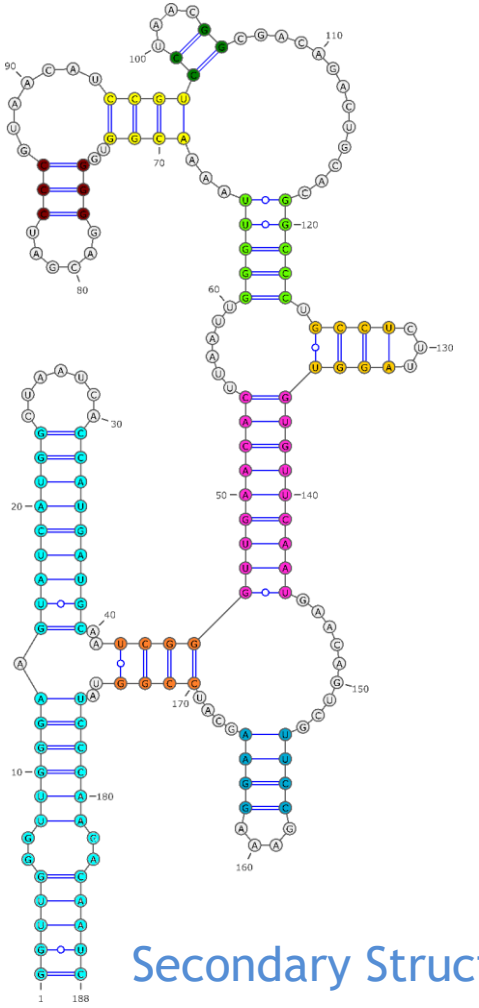


In collaboration with Y.Ponty's group we developed IPANEMAP a pipeline integrating multiple sets of probing data (Saaidi et al. NAR 2020)

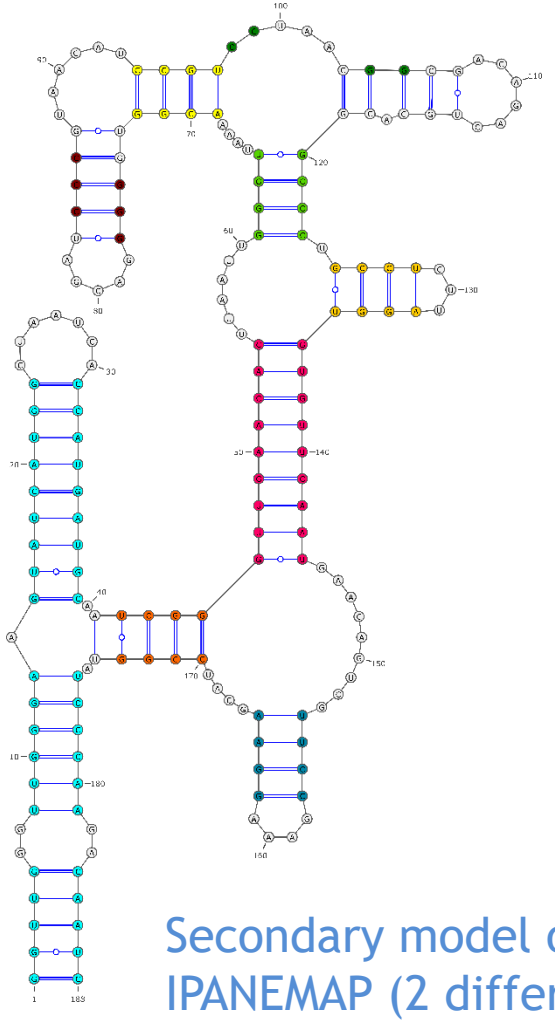
… New release including phylogenetic data soon

cnrs Université Paris Cité

IPANEMAP prediction using two probes



Secondary Structure



Secondary model obtained with IPANEMAP (2 different probes)

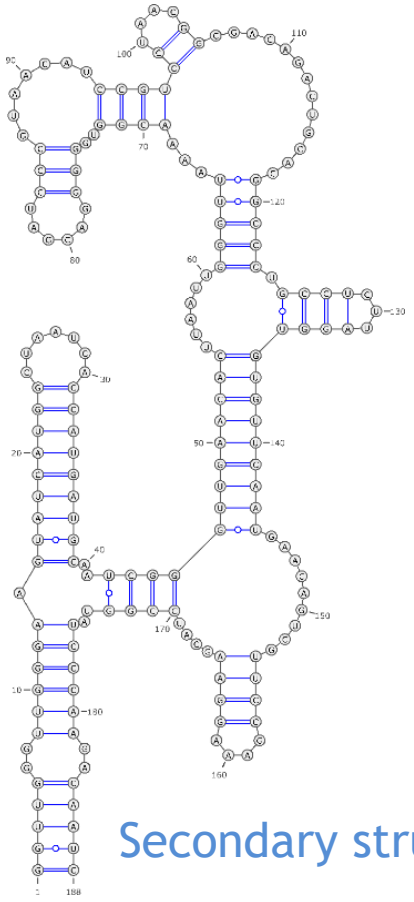
Some improvment

…

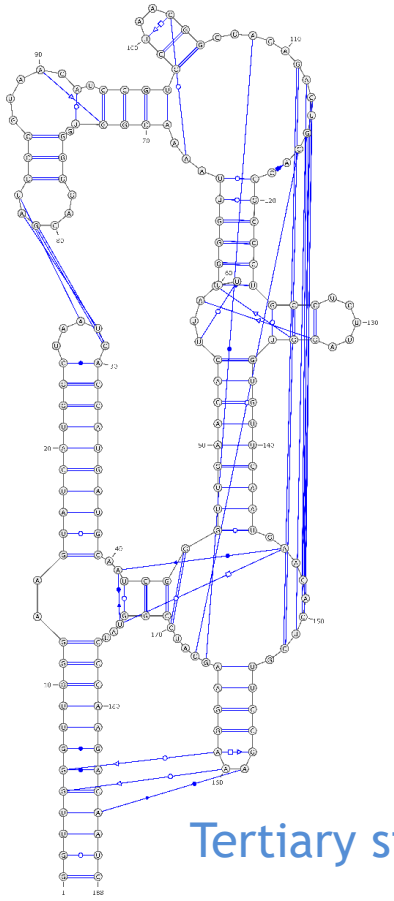
But still not perfect

cnrs Université Paris Cité

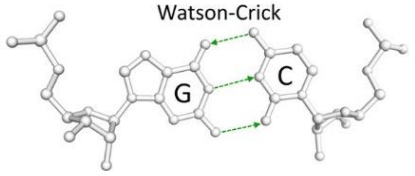
Secondary vs tertiary structure

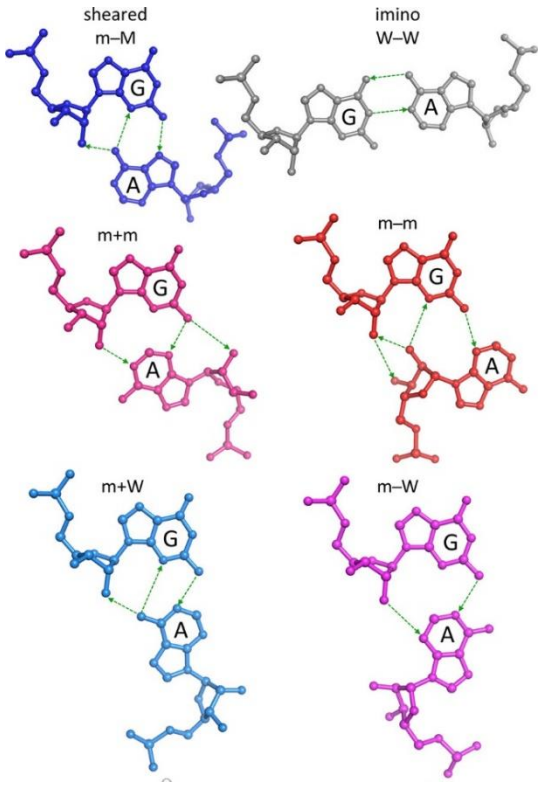


Secondary structure



Tertiary structure



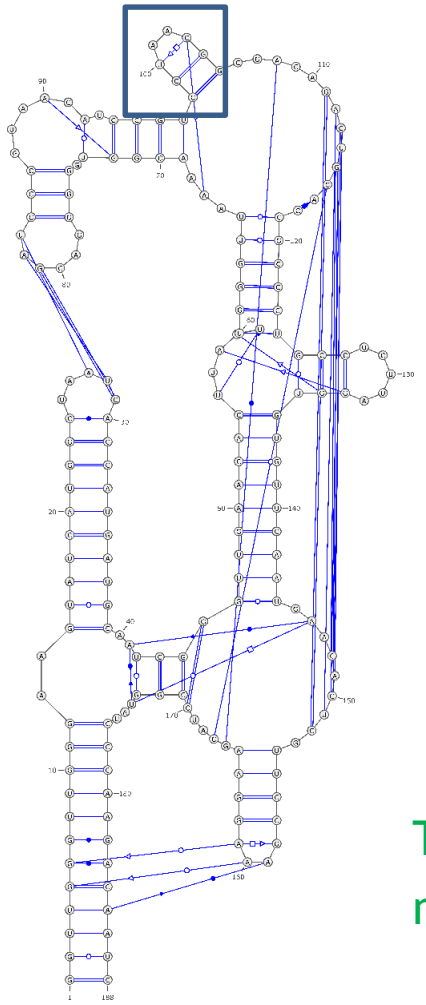


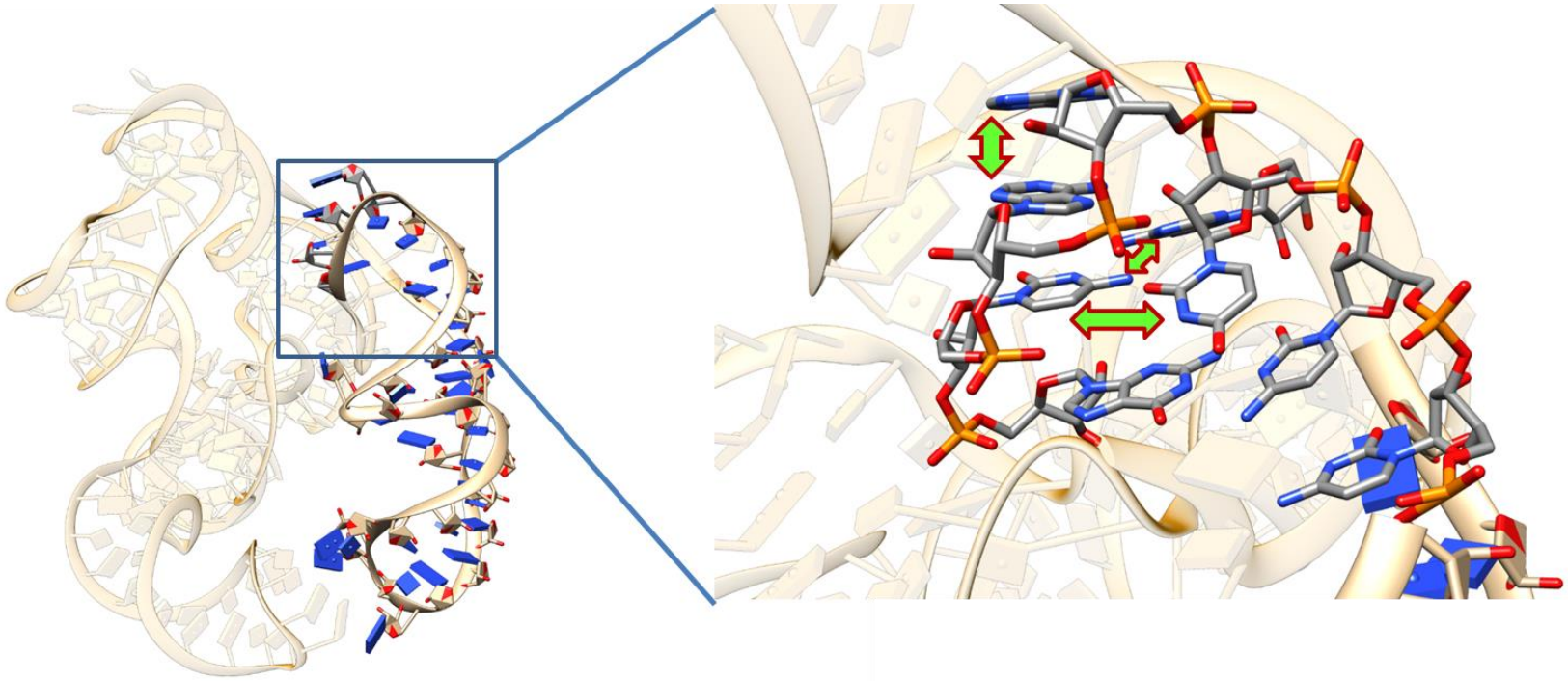
31 non canonical pairings
12 families

Interactions such as pseudoknots or non-canonical pairings can not be taken into account because of the combinatory involved

cnrs Université Paris Cité

Structure stabilisation





Tertiary structure brings much energy that is not captured by the energy models driving the modeling software leading to spurious predictions

⇨ Detect nucleotides involved and exclude them from the secondary structure modeling

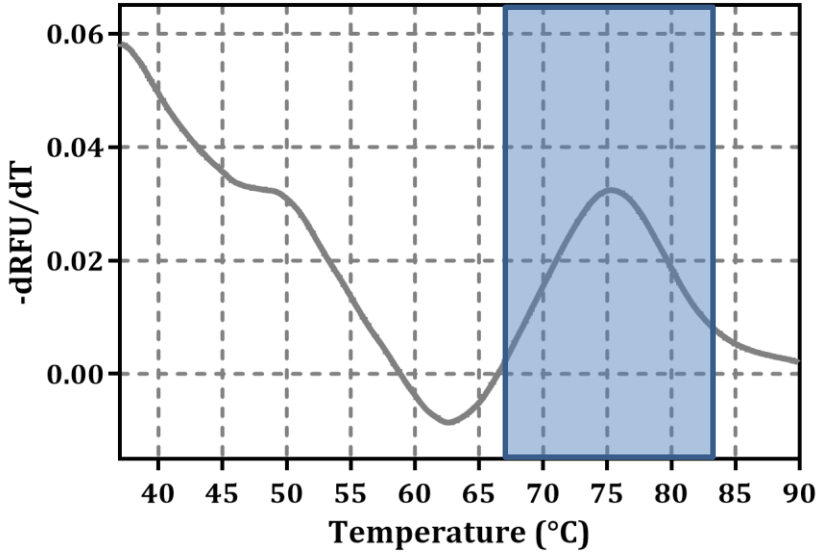
cnrs Université Paris Cité

Thermal unfolding of the RNA structure

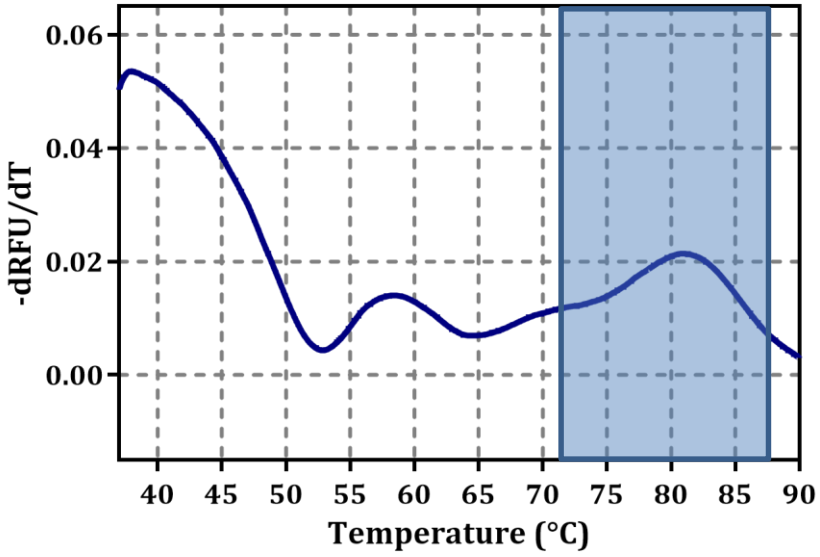
RNA was incubated with a probe fluorescent when bound to single strand region :

- Fluorescence increases when the structure unfolds (more single strand available)
- Fluorescence decreases with temperature (affinity fo the probe decreases)

The experiment was carried out in abscence or presence of Mg^{2+} which stabilises tertiary interactions



In absence of Mg2+

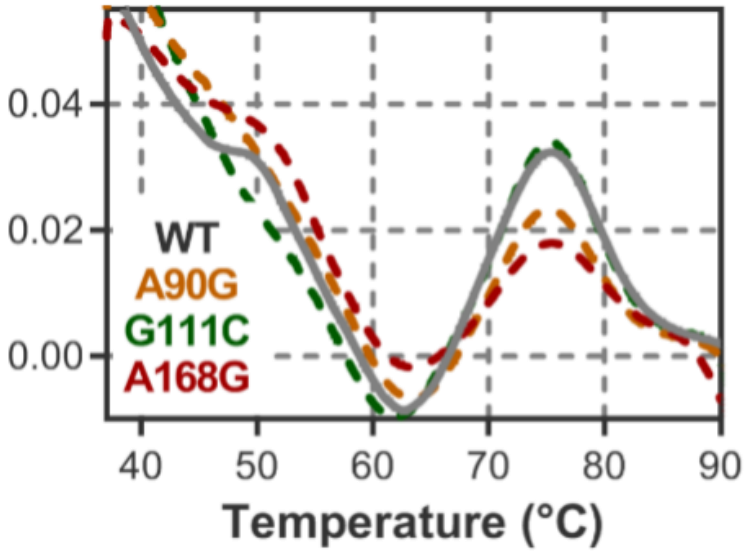


In presence of Mg2+

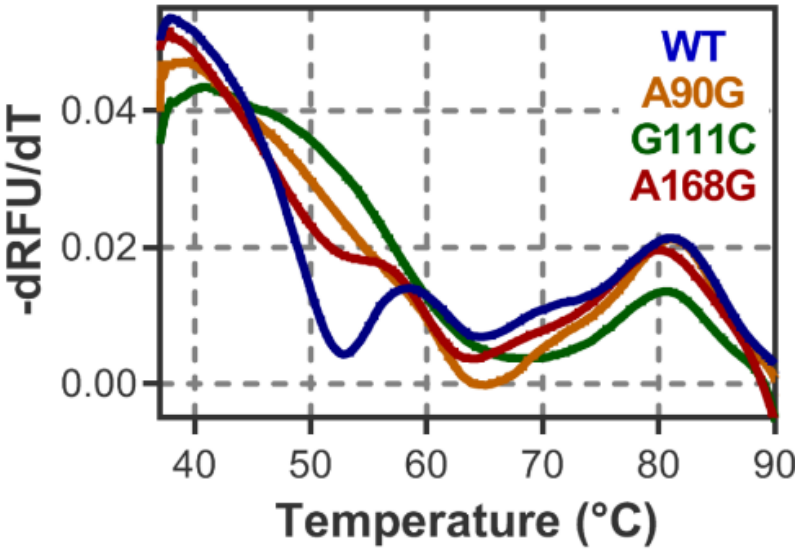
cnrs Université Paris Cité

Thermal unfolding of the RNA structure

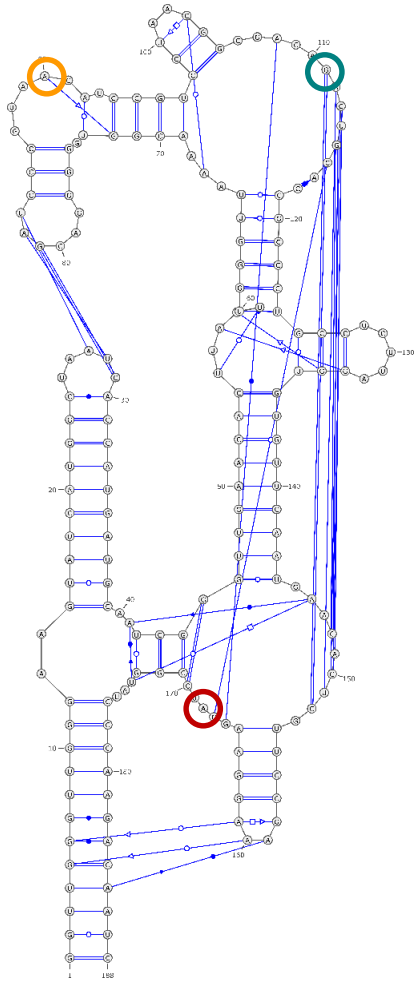
Mutants disrupting different tertiary contacts were submitted to UV-melting experiments



In absence of Mg2+



In presence of Mg2+

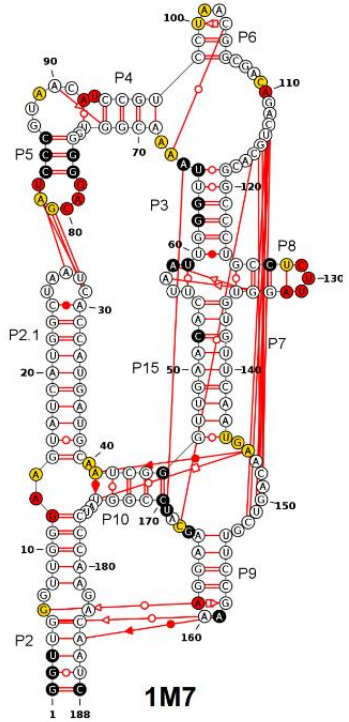


1. Tertiary structure unfolds cooperatively at a lower temperature than secondary structure
2. Tertiary structure is unstable in the absence of Mg^{2+}

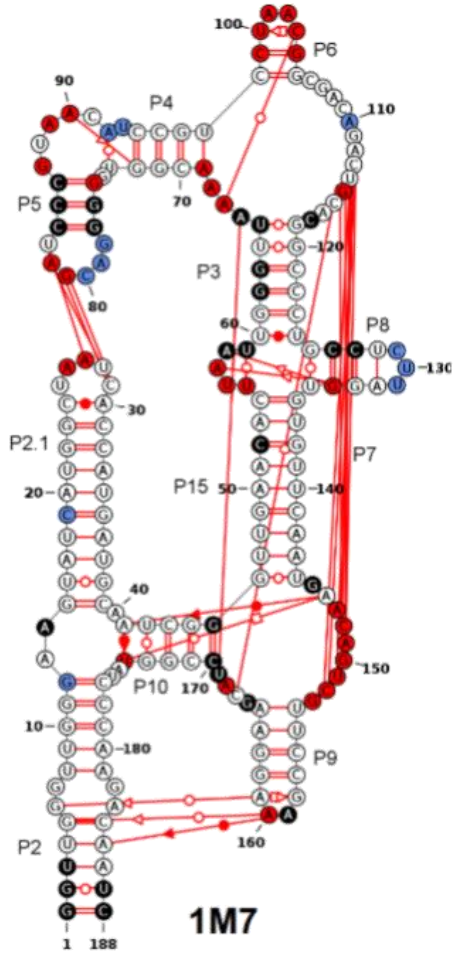
cnrs Université Paris Cité

Divalent cations stabilise the tertiary structure

Di-LcRz was submitted to SHAPE structure probing in absence of presence (5 mM) of $MgCl_2$



Mg^{2+}



1. Mg^{2+} Differential probing capture most of the nucleotides involved in the tertiary structure … But not all of them
2. The 5'part of the Pseudoknot is involved in an alternative structure in absence of Mg^{2+}

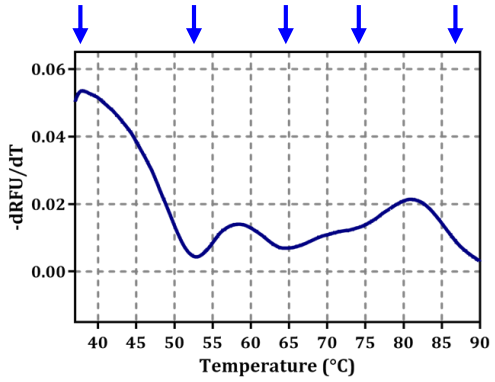
Differential map Mg^{2+} VS no Mg^{2+}

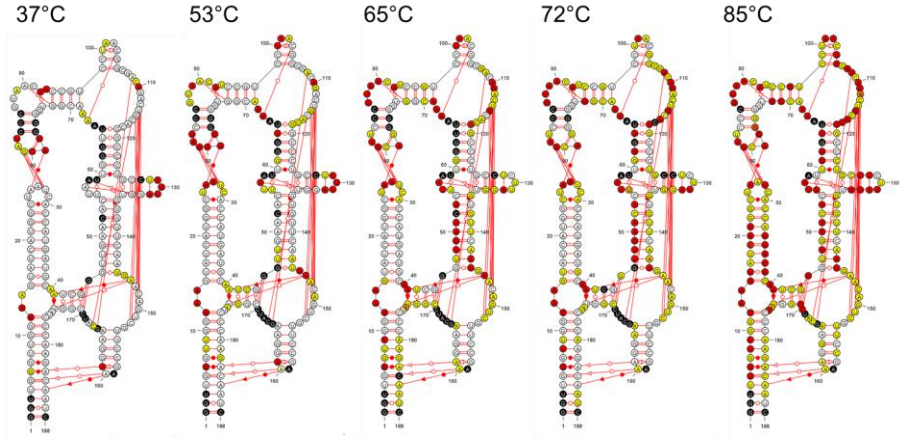
- *More reactive in absence of Mg2+*
- *Less reactive in absence of Mg2+*

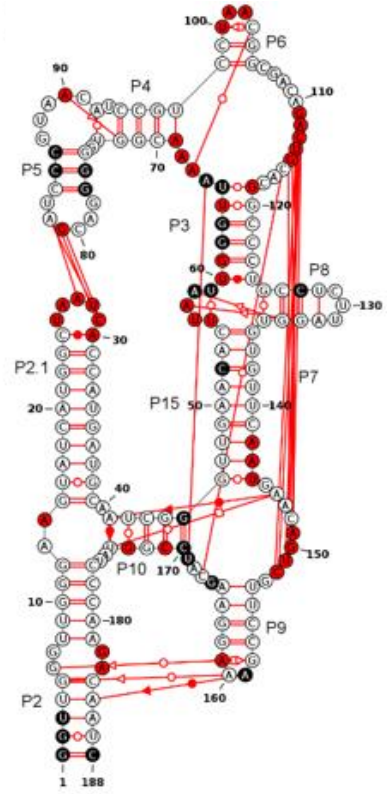
cnrs Université Paris Cité

Following thermal denaturation with shape

Di-LcRz was submitted to SHAPE structure probing at increasing temperature







37°C-53°C Differential map

- *More reactive at 53°C*

37°C-53°C diferential shape efficiently detects nucleotides involved in tertiary structure

Université Paris Cité

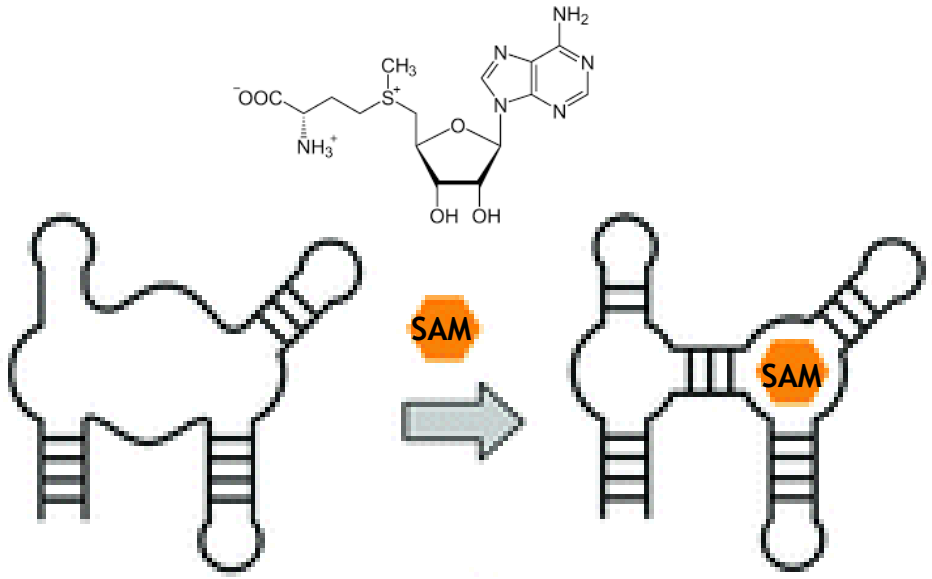
Predicting tertiary interactions

1. In order to detect non-canonical pairings, and pseudoknots RNA are probed in different conditions (temperature, ionic conditions …). The differential reactivity map can be used to inform the prediction software to improve the modeling (under development)

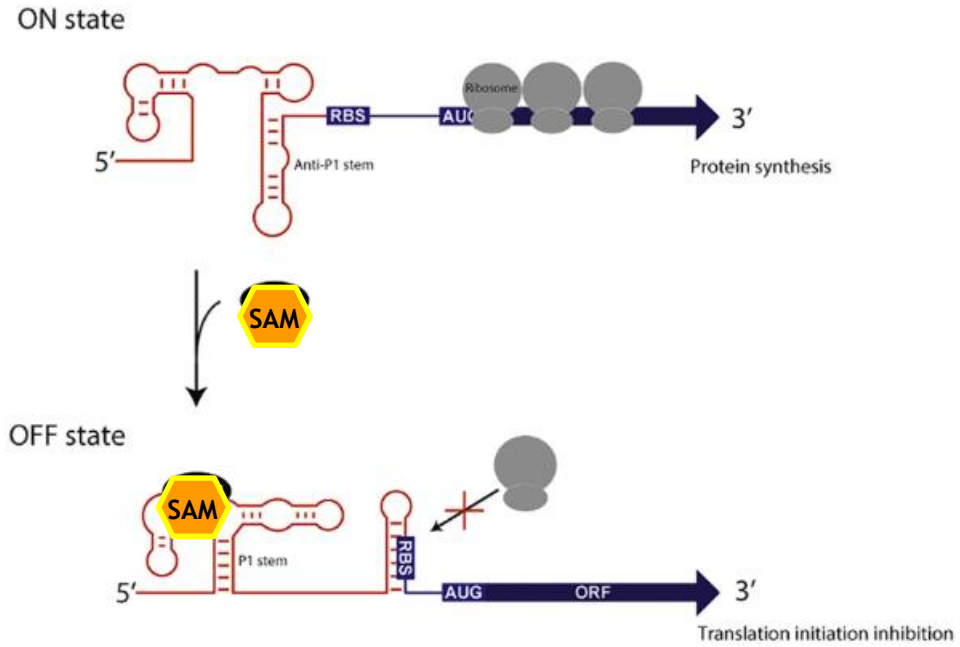
2. To better understand the evolutionary constraints on RNA structure and structural dynamics. We applied such technology to study artificial and natural RNA aptamers

RNA that bind a small molecule - SAM

S-Adenosyl Methionine (SAM)



SAM binding RNA aptamers

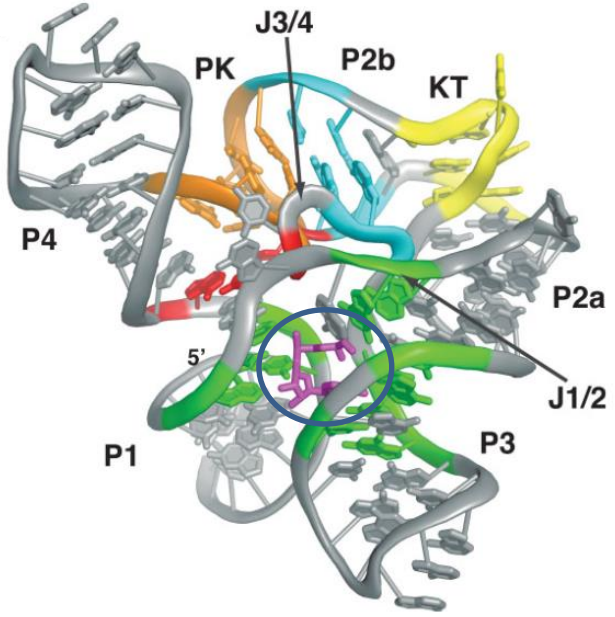


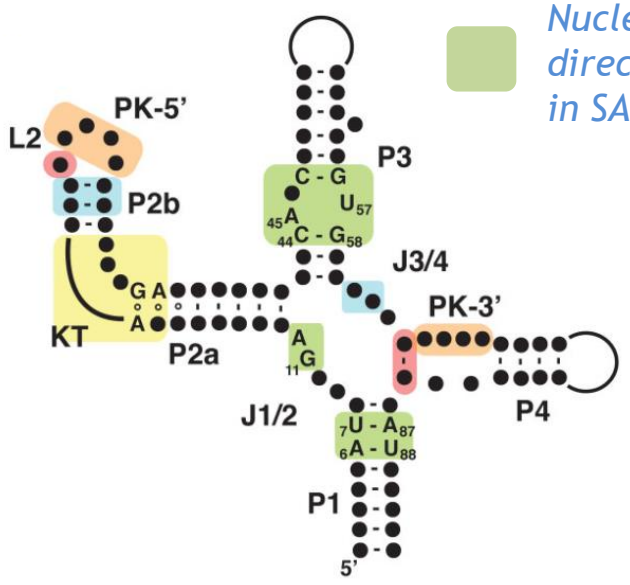
SAM aptamers are part of riboswitches

1. Model to study structural dynamics
2. Design artificial aptamers using machine learning to better understand the evolutionary constraints on these sequences

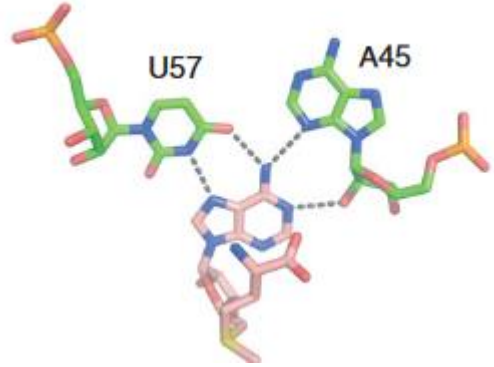
cnrs Université Paris Cité

Molecular basis of SAM recognition





Nucleotides directly involved in SAM binding

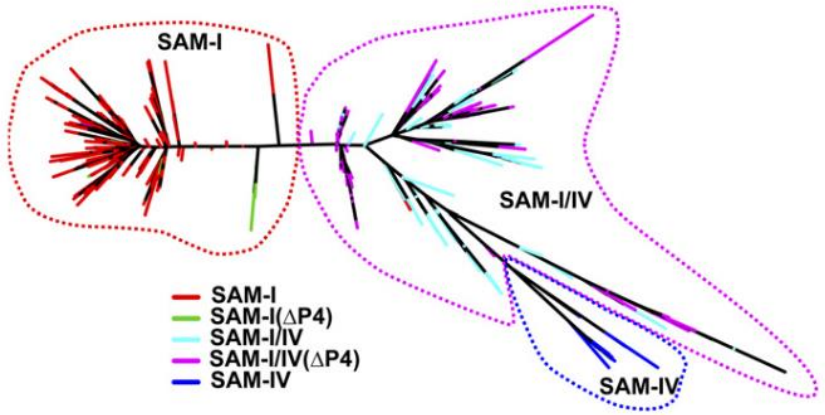


From Batey, 2011

1. SHAPE probing pattern will reflect the structure
2. SAM binding will protect specific nucleotides from SHAPE reactivity

cnrs Université Paris Cité

Natural sequence diversity



SAM aptamers are present in **many bacteria**

RFAM, lists 6161 sequences of natural bacterial SAM aptamers, among these

1. 457 are « cured » sequences (the « seed » set). 200 sequences representatives of the diversity were sampled to be studied by SHAPE

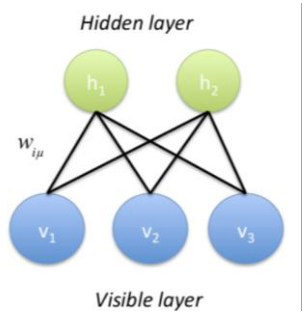
2. 5704 were identified in sequence databases by sequence/structure homology (the « full » set) . 55 sequences representatives of the diversity were sampled to be studied by SHAPE

cnrs Université Paris Cité

Designing artificial aptamers by machine learning

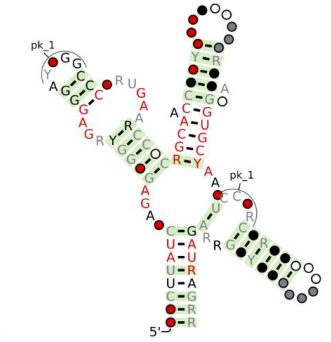
457 « cured» (seed) sequences of RFAM were used to design artificial aptamers

- Restrained Boltzmann machine, a simple two layers neural network was trained with the 457 SAM aptamer sequences to generate new artificial aptamers



Statistics/probability only

- « INFERNAL » a software based on covariation model, was used to design new artificial aptamers



Based on RNA structure knowledge

cnrs Université Paris Cité

Experimental procedure

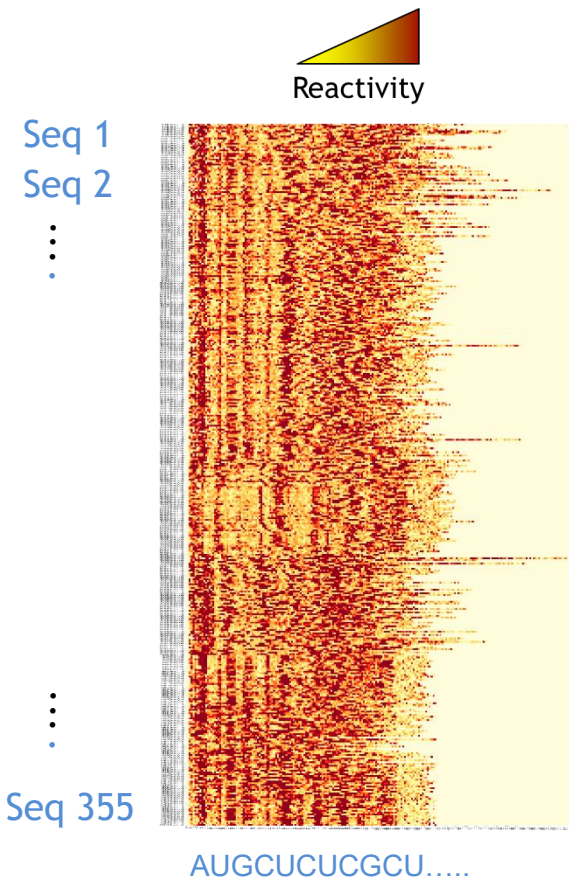
355 sequences (18 infernal, 82 rbm, 200 seed, 55 full) were chemically synthetized as DNA oligonucleotides

355 RNA sequences

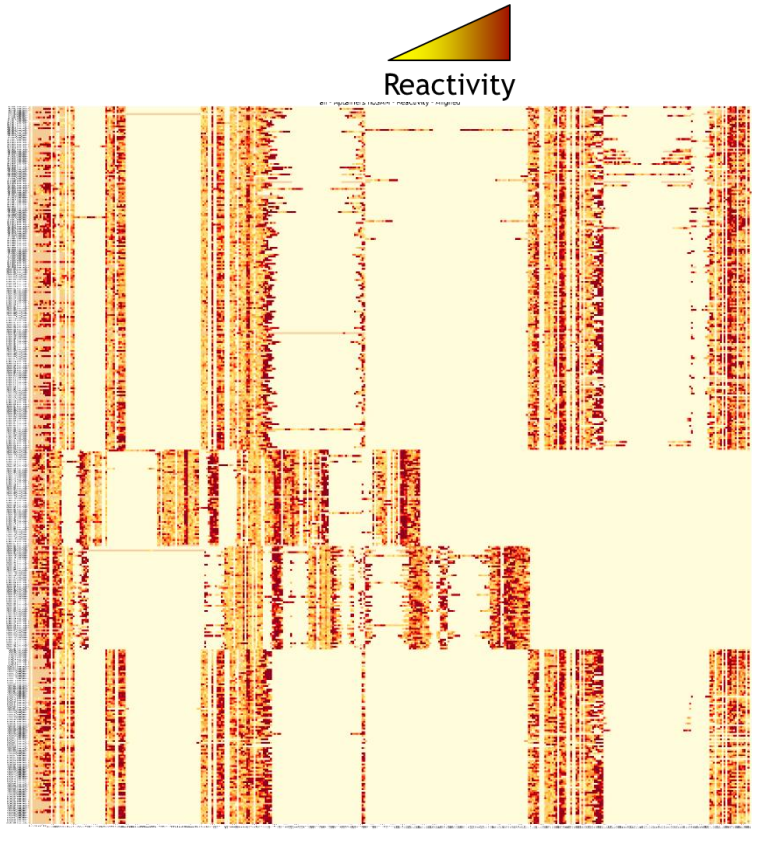
SHAPE probing in various conditions
- +/- SAM
- +/- Mg^{2+}
- elevated temperature

Differential maps for 355 sequences

Probing RNA structure



Alignment
Clustering



Type I

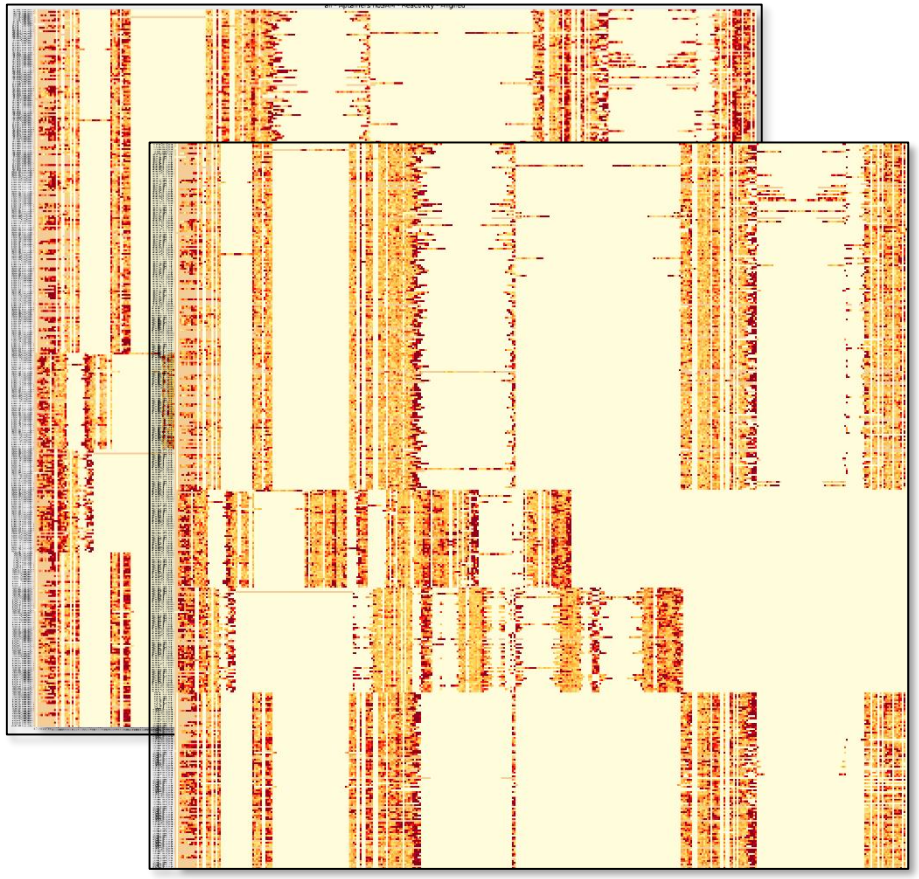
Type IV

Type I/IV

Artificially designed

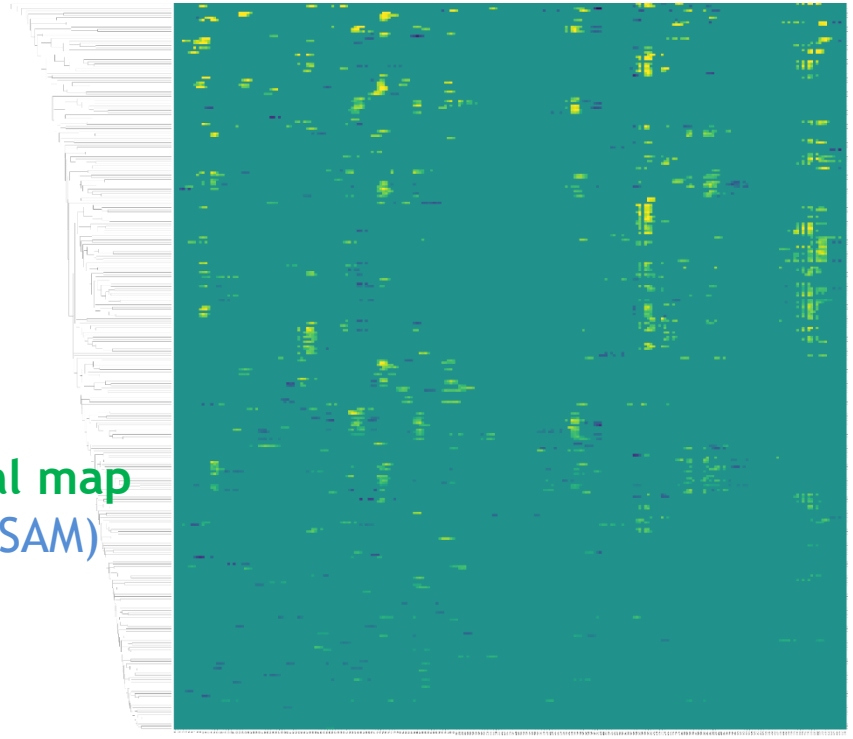
Differential map

without SAM



with SAM

Differential map
(SAM – noSAM)



Reactivity in presence of the ligand

- equal
- reduced
- enhanced

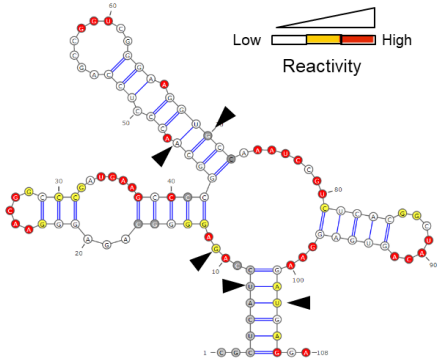
Interaction is unambiguously detected and characterized

cnrs Université Paris Cité

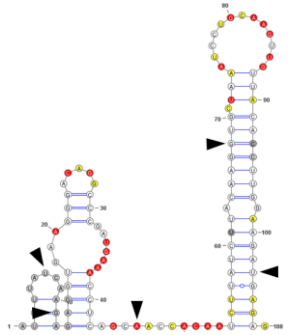
Target binding

All « seeds » bind SAM while only half of the « full » sequences of RFAM are bona fide aptamers

1/3 of the artificial sequences generated by RBM machine learning are bona fide aptamers while none of those selected with INFERNAL bind SAM



1/3 of RBM adopt the correct secondary structure but do not form a complex with SAM



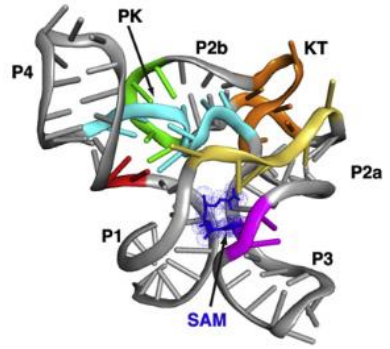
1/3 of RBM and all infernal sequences adopt an alternative folding

cnrs Université Paris Cité

Structural dynamics of SAM aptamers

Differential maps obtained in presence or absence of Mg^{2+} and SAM allows for distinguising **different classes** of aptamers:

1. The **tertiary structure is not stable** in the absence of SAM – the RNA folds around the ligand
2. The binding site **pre-exist to the ligand** presence



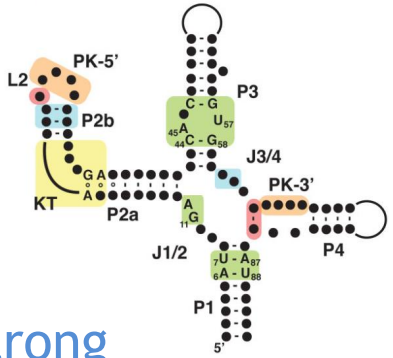
Such different structural dynamics will influence the switching **rate of the cognate riboswitch**

- Transcriptional vs translational switch?
- Related to the biotope?
- Related to bacterial growth?

cnrs Université Paris Cité

Conclusions

1. Designing an RNA on the basis of a consensus sequence/structure is not sufficient to obtain an RNA actually adopting the desired structure because such process:
   - Does not counter select for alternative secondary structure folding
   - Does not take into account the tertiary structure

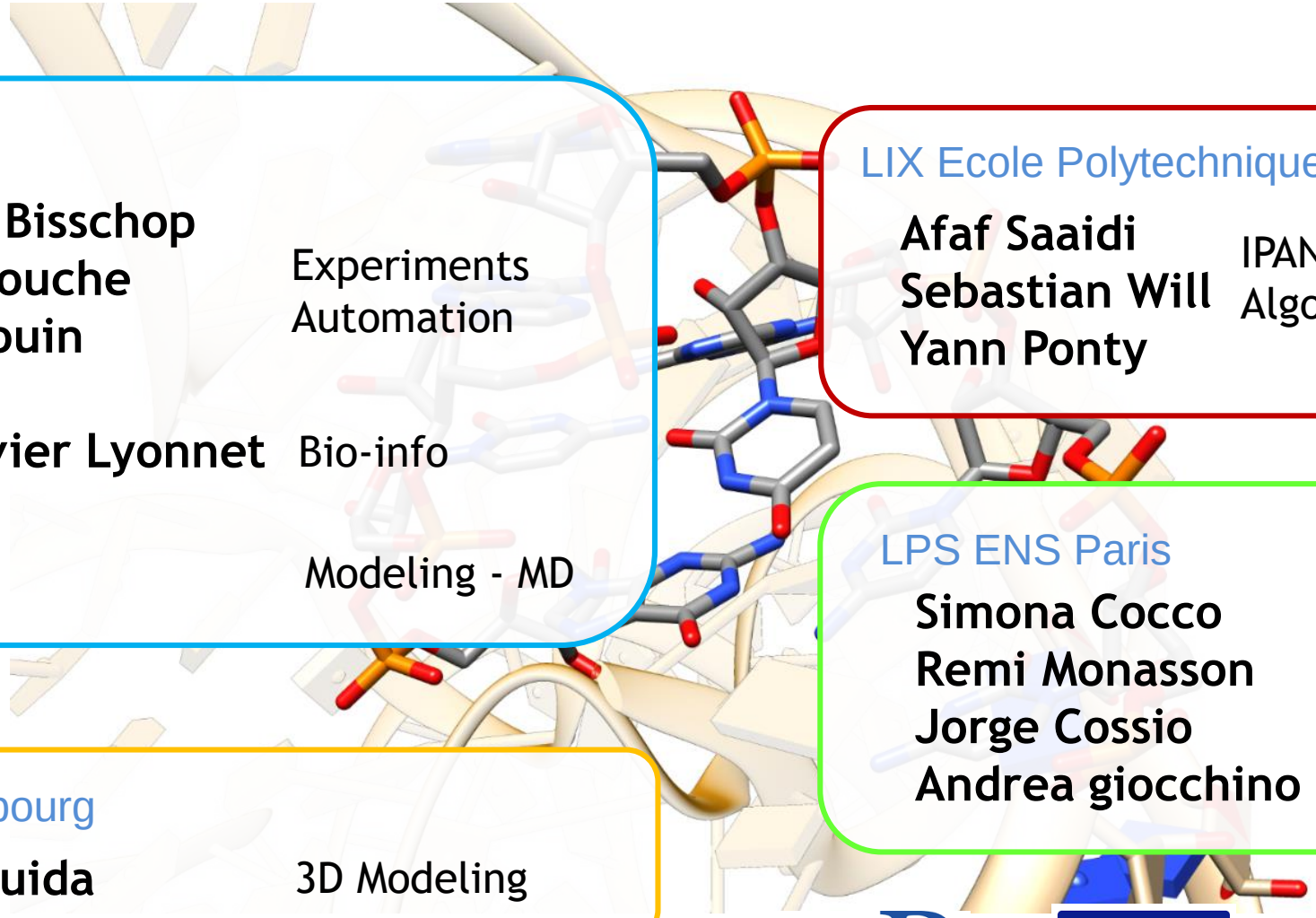


2. The « negative folding » constraint captured by statistical methods is a strong evolutionary constraint

3. Different RNA sequence with strong structural homology can have different dynamics influencing their function

cnrs Université Paris Cité

Team and collaborators



CitCOM Paris

Gregoire de Bisschop **Delphine Allouche** **Pierre Hardouin**	Experiments Automation
Francois Xavier Lyonnet	Bio-info
Elisa Frezza	Modeling - MD

LIX Ecole Polytechnique Palaiseau

Afaf Saaidi **Sebastian Will** **Yann Ponty**	IPANEMAP Algorithm and programming

LPS ENS Paris

Simona Cocco **Remi Monasson** **Jorge Cossio** **Andrea giocchino**	Machine learning RNA design

GMGM Strasbourg

Benoit Masquida	3D Modeling

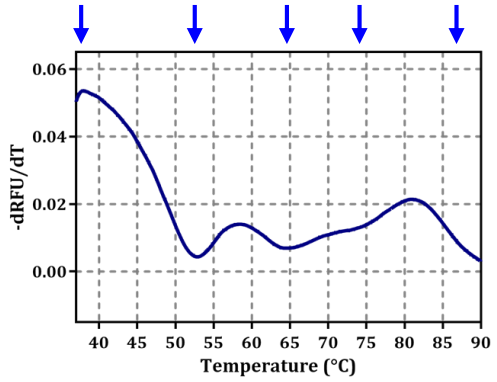
AGENCE NATIONALE DE LA RECHERCHE ANR

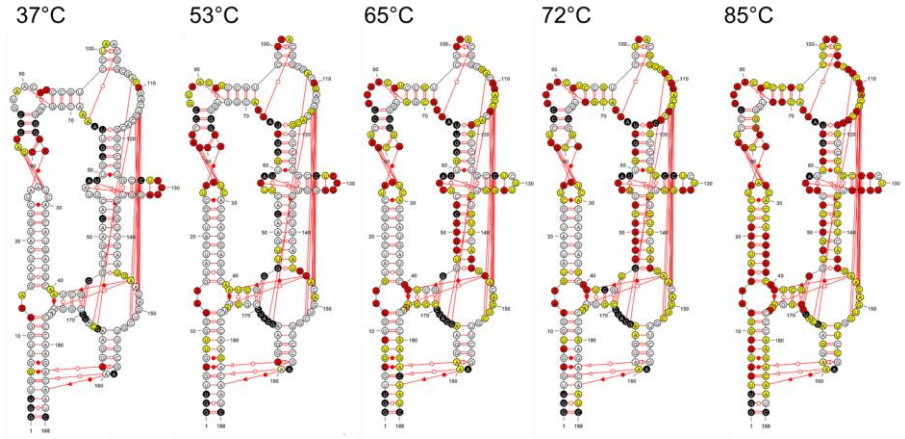
FRM FONDATION RECHERCHE MÉDICALE

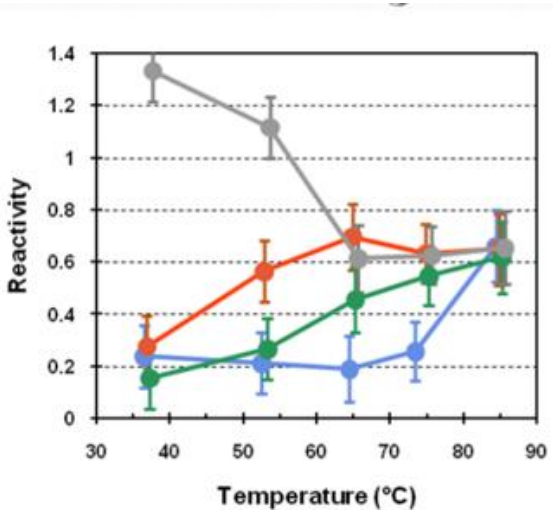
cnrs Université Paris Cité

Following thermal denaturation with shape

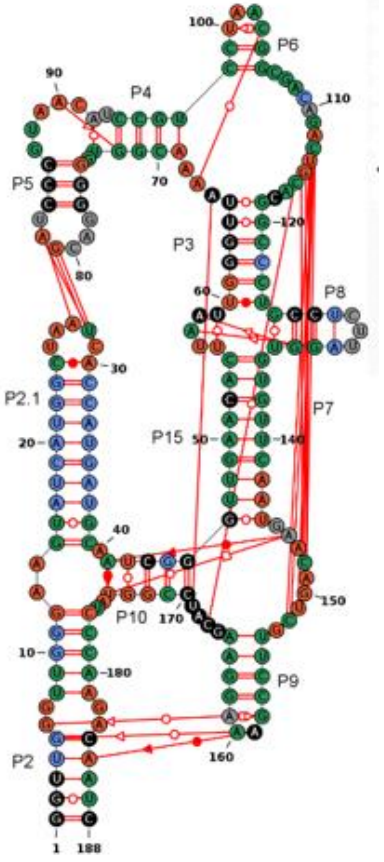
Di-LcRz was submitted to SHAPE structure probing at increasing temperature







K-mean clustering of the nucleotide according to the evolution of their shape reactivity in function of temperature

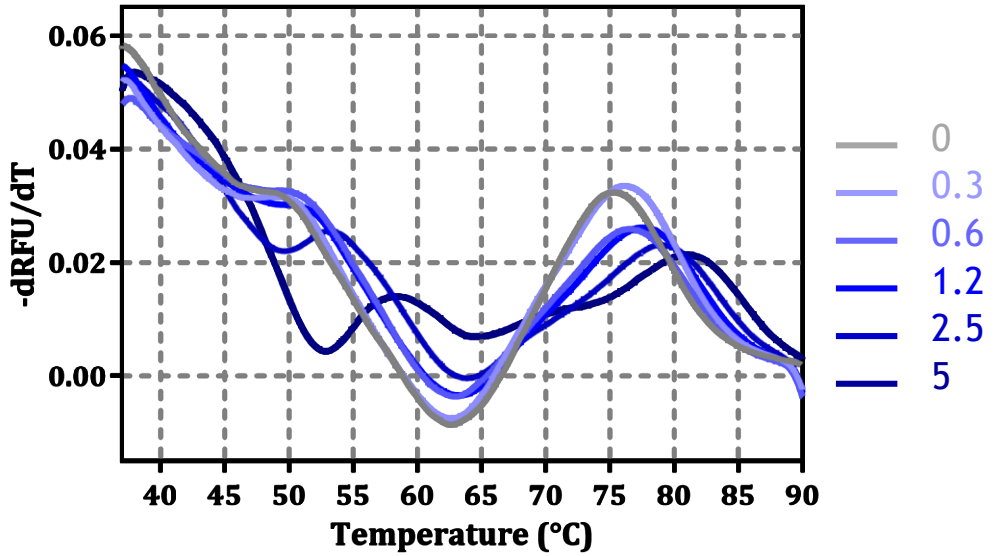


Diferential thermal shape efficiently detects nucleotides involved in tertiary structure

cnrs Université Paris Cité

Thermal unfolding of the RNA structure

WT RNA was submitted to UV metling experiments in the presence of increasing concentrations of Mg2+



1. Mg^{2+} stabilizes RNA tertiary structure
2. Di-LcRz was submitted to SHAPE structure probing in absence or presence (5 mM) of $MgCl_2$

cnrs Université Paris Cité