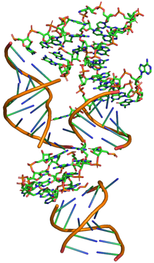
Arbitrarily large RNA structural modules

Vladimir Reinharz

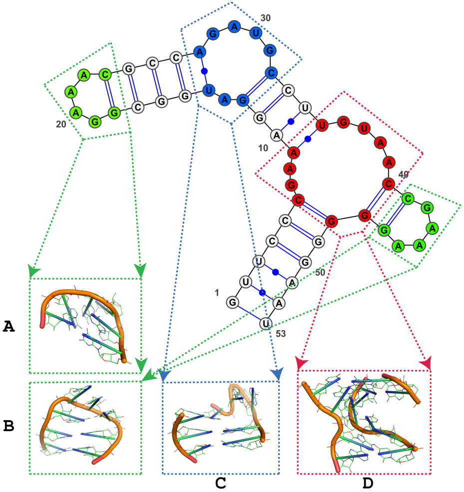
UQAM

`cbe.uqam.ca`

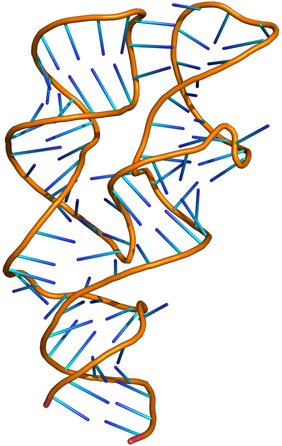


Benasque 2022

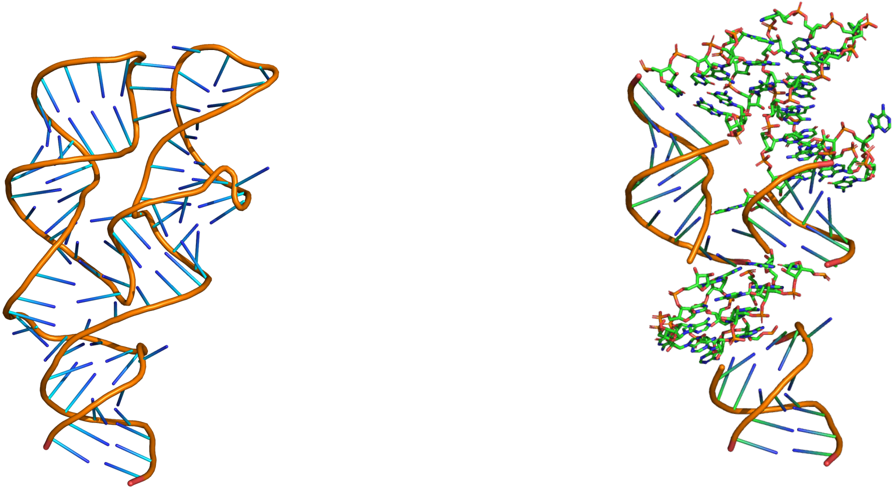
Loops are highly organized in 3D



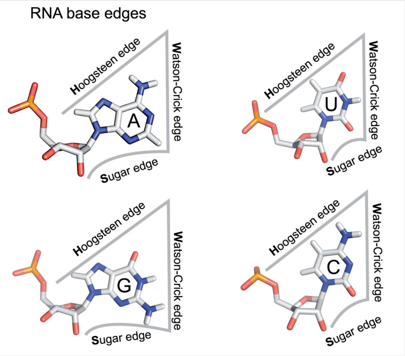
RNA structure



RNA structure



Geometries of nucleotides interactions
Leontis–Westhof annotations



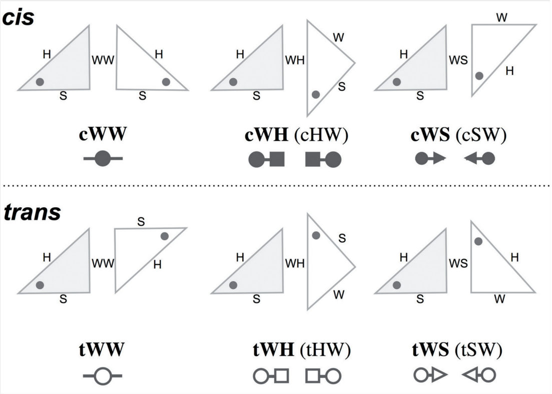
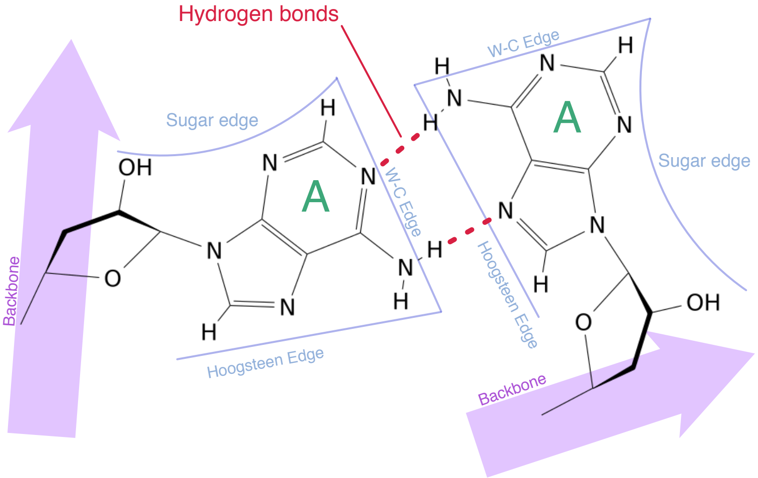
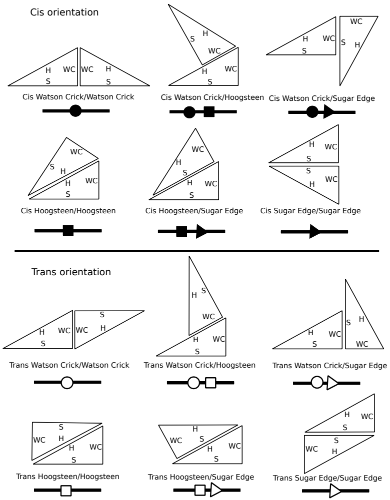


Figure adapted from
Almakarem et al, 2011

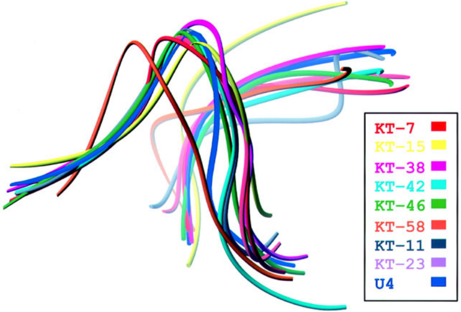
Geometries of nucleotides interactions
Leontis–Westhof annotations

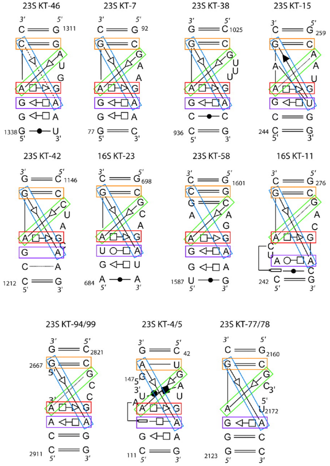


Geometries of nucleotides interactions
Leontis–Westhof annotations



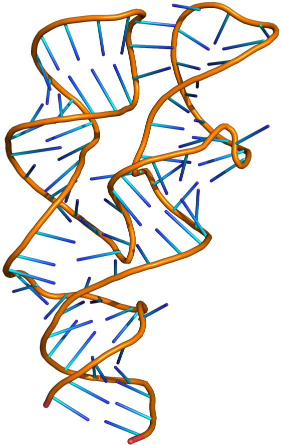
Kink turn
local structural motif



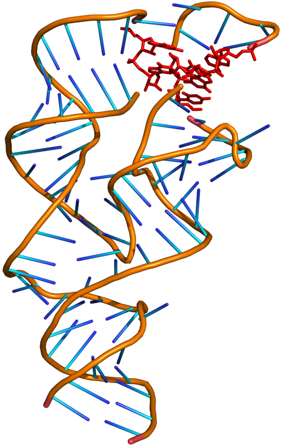


Klein et.al. 2001, Lescoute et.al. 2005

Loop-loop interactions



Loop-loop interactions



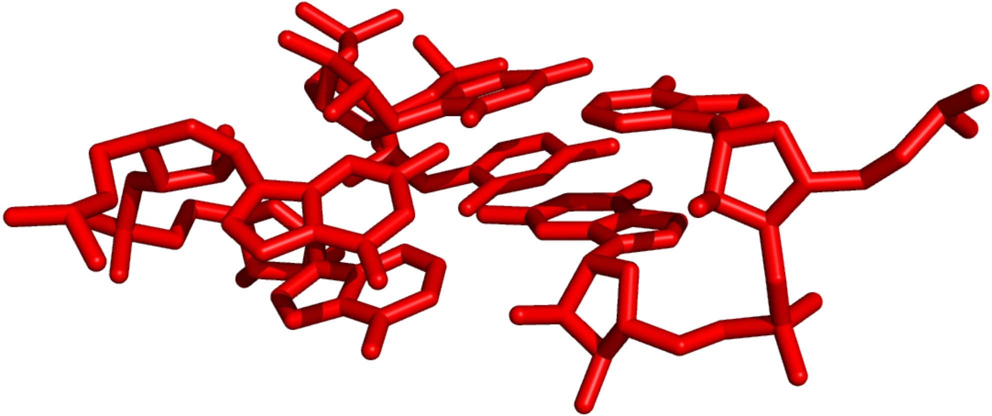
Loop-loop interactions

Mining for recurrent long-range interactions in RNA structures reveals embedded hierarchies in network families

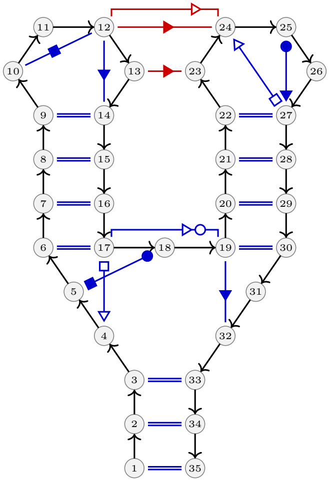
Vladimir Reinharz, Antoine Soulé, Eric Westhof, Jérôme Waldispühl, Alain Denise

Nucleic Acids Research, Volume 46, Issue 8, 4 May 2018, Pages 3841–3851, https://doi.org/10.1093/nar/gky197

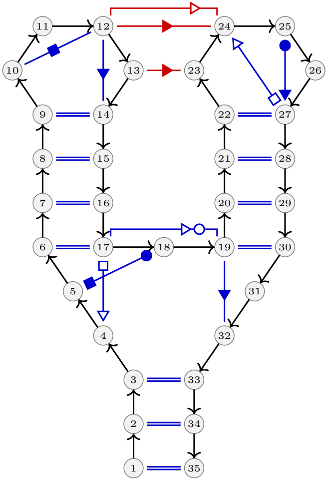
Loop-loop interactions

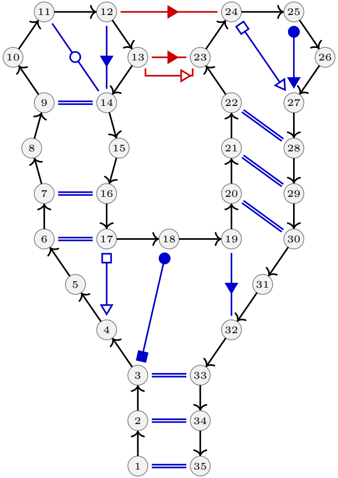


RNAs as graphs



Are sub-structures conserved?





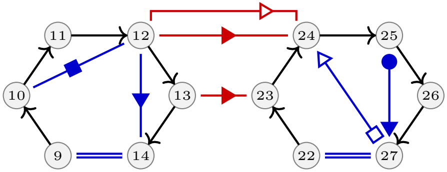
Subgraphs isomorphisms
Are two subgraphs similar?

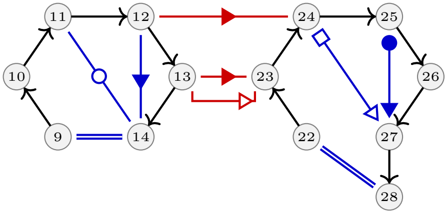
- In general NP-hard
- Can have exponential number of solutions
  (in number of nucleotides)

Subgraphs isomorphisms
Are two subgraphs similar?

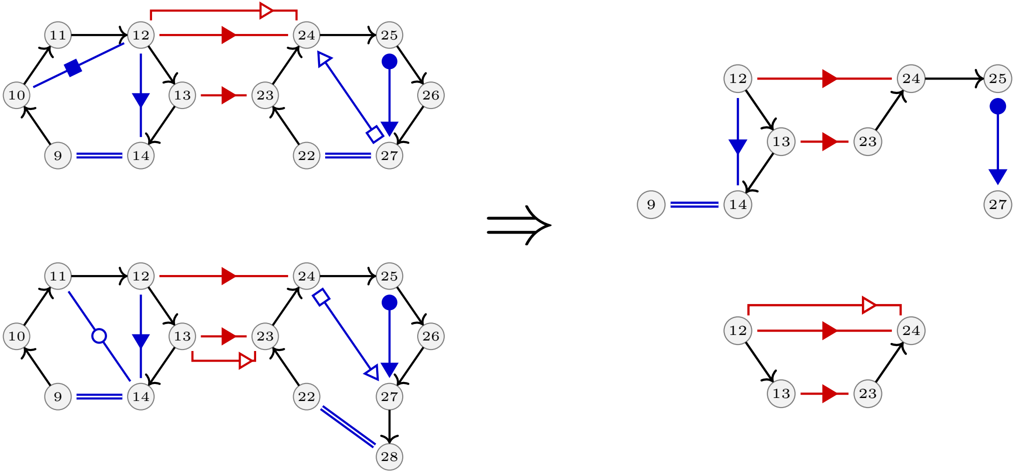
- In general NP-hard
- Can have exponential number of solutions (in number of nucleotides)
- Take advantage of biological constraints
- Reduce problem to manageable size

Find identical subgraphs



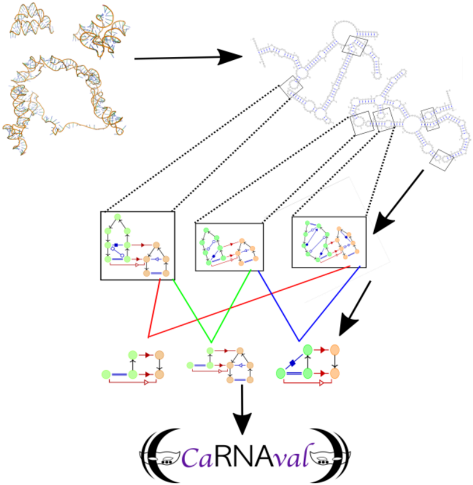


Find identical subgraphs



All structures analyzed (in 2018)

`carnaval.lri.fr`

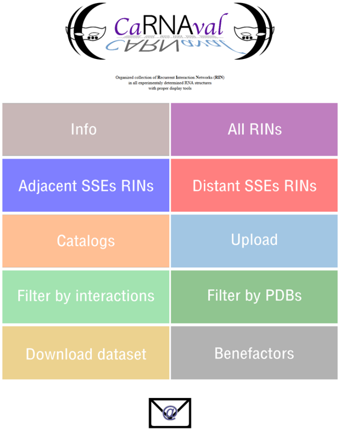


CaRNAval

carnaval.lri.fr

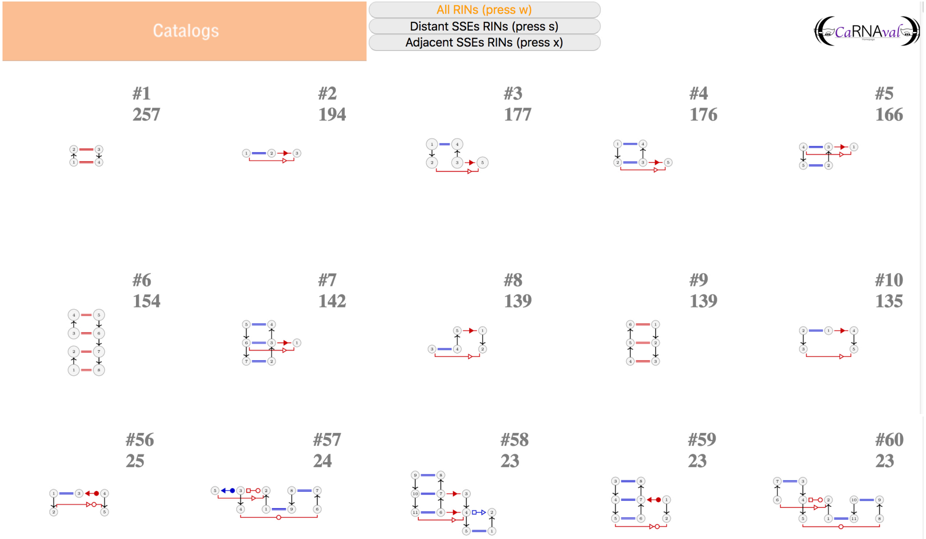
845 RNAs with 1426 pairs of loops

337 conserved sub-structures, 6056 instances



Organized collection of **R**ecurrent **I**nteraction **N**etworks (**RIN**)
in all experimentaly determined RNA structures
with proper display tools

Catalogue



#1
257

#2
194

#3
177

#4
176

#5
166

#6
154

#7
142

#8
139

#9
139

#10
135

#56
25

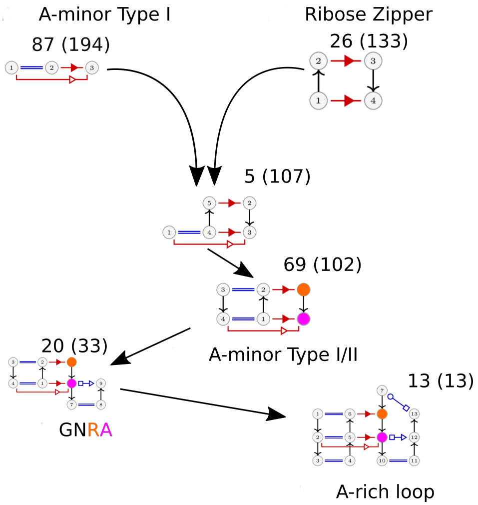
#57
24

#58
23

#59
23

#60
23

A-minor and Inheritance



A-minor Type I

87 (194)

Ribose Zipper

26 (133)

5 (107)

69 (102)

A-minor Type I/II

20 (33)

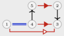
GNRA

13 (13)

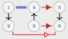
A-rich loop

A-minor and Inheritance

#15
107

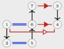


#16
103

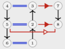


#17
102

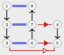
#24
80



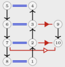
#25
79



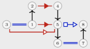
#30
63



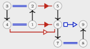
#36
43



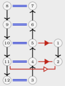
#43
34



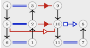
#44
33



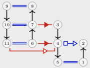
#48
31



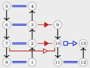
#50
29



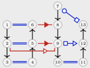
#58
23



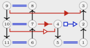
#62
20



#74
13

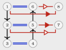


#87
11

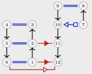


A-minor and Inheritance

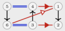
#99
9



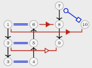
#102
8



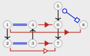
#107
8



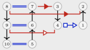
#108
8



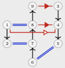
#109
8



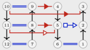
#110
8



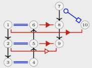
#117
7



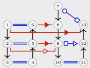
#118
7



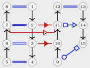
#127
7



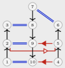
#134
6



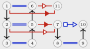
#137
6



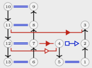
#138
6



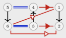
#150
6



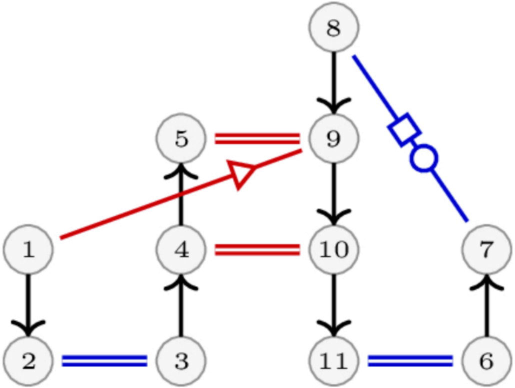
#162
5

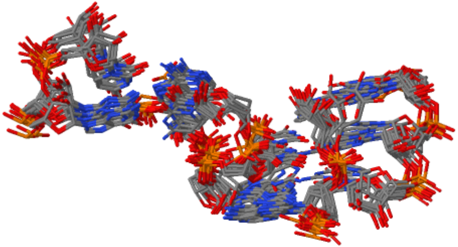


#165
5

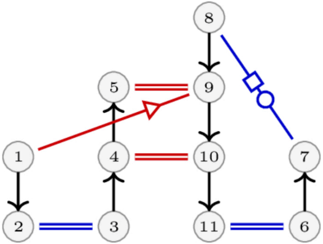


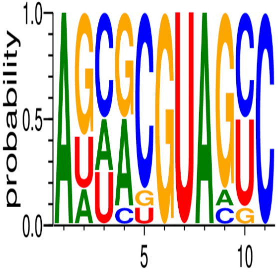
Graph conservation → structure conservation
Found in ribosomes, ribozymes and riboswitches

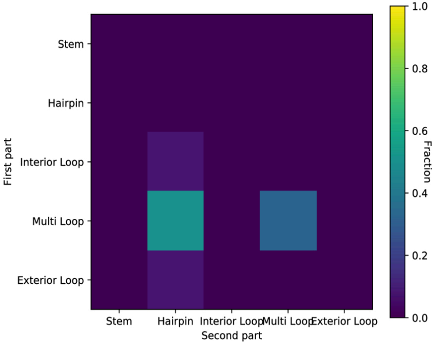




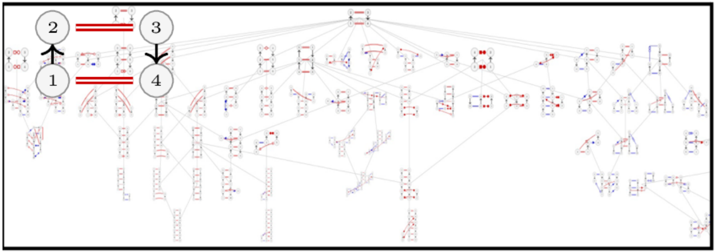
RIN 78

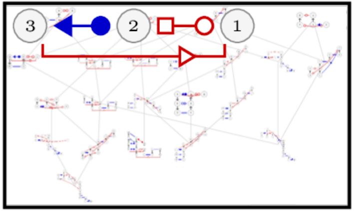

8
5
9
1
4
10
7
2
3
11
6

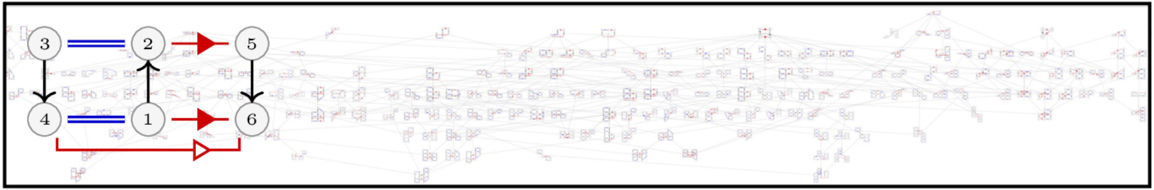

probability
1.0
0.5
0.0
5
10

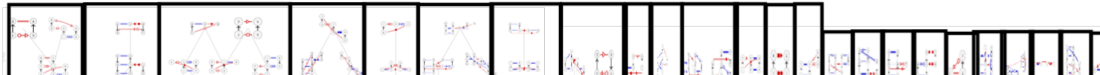

Stem
Hairpin
Interior Loop
Multi Loop
Exterior Loop
First part
Stem
Hairpin
Interior Loop
Multi Loop
Exterior Loop
Second part
1.0
0.8
0.6
0.4
0.2
0.0
Fraction

All RINs networks







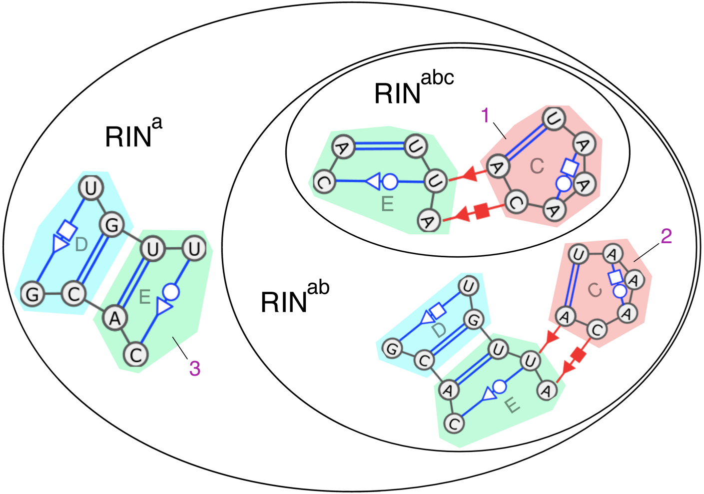


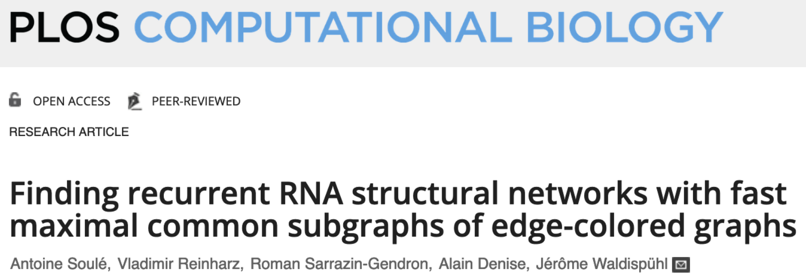
CaRNAval 1 — Conclusion and drawbacks

- First fully automated method retrieving and clustering RNA recurrent interaction networks with long-range interactions
- Known structures recovered and new ones observed
- Map of the modular network of RINs : inclusion relations, combination

- Limited to pairs of loops in the same RNA

Relaxing constraints





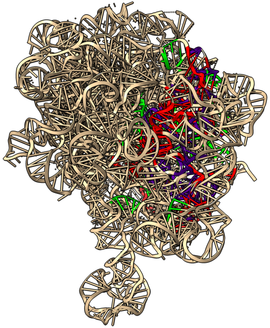
PLOS COMPUTATIONAL BIOLOGY

OPEN ACCESS PEER-REVIEWED

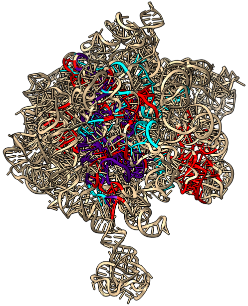
RESEARCH ARTICLE

Finding recurrent RNA structural networks with fast maximal common subgraphs of edge-colored graphs

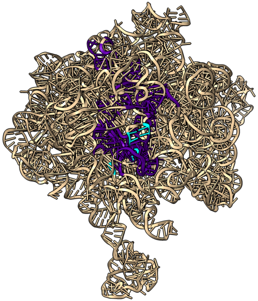
Antoine Soulé, Vladimir Reinharz, Roman Sarrazin-Gendron, Alain Denise, Jérôme Waldispühl



T. Thermophilus 70S

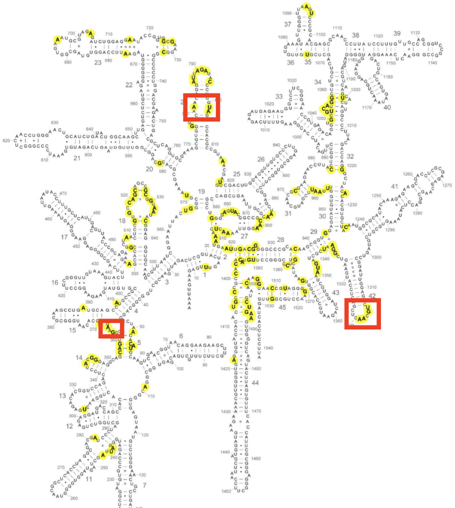


E. Coli 70S



P. Aeruginosa 50S

ToDo: Perfect sequence conservation in 16S
Noller, Donohue, and Gutell, RNA, 2022



A last level of complexity

RNAs epigenetic, impact of chemical modifications

- Necessary for ribosome to recruit some amino acids (Grosjean & Westhof, NAR, 2016)
- Boost gene expression and restrict differentiation (Garber, Nat., 2019)
- Can destabilize or stabilize structure (Tanzer et al., Methods, 2018)
- Involved in human diseases (Asano et al., NAR, 2018)
- Involved in cancer (Chi, Nat., 2017 & Deng et al., Cell research, 2018)
- New drugs as interfering-RNAs (Kim et al., Nat. Rev. Gen., 2007)
- Viral-less in vivo gene editing (Yin et al., Nat. Biotech., 2017)

RNAs chemical modifications

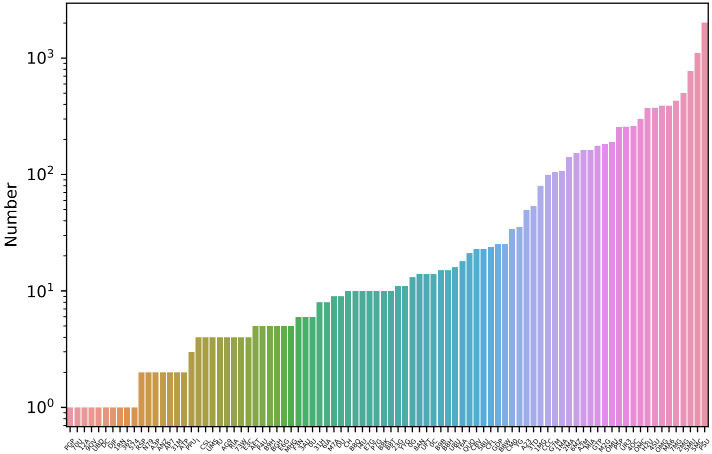
What about the structure?

In all 151 079 3D complexes in pdb.org

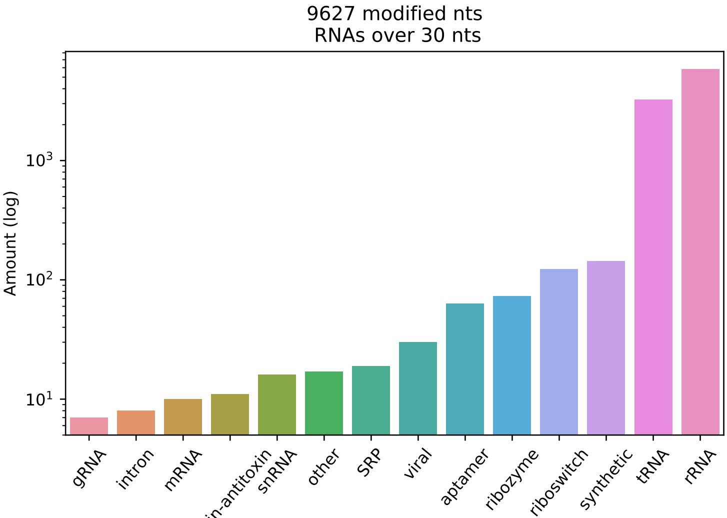
- 4324 complex of molecules with RNA chains in known 3D structures
- 4208 have at least one chemical modification
- In total 190 different chemical modifications
- 4 855 855 nucleotides
- 11 682 chemically modified

Distribution of chemical modifications known 3D structures

90 modifications, 9627 modified nts
RNAs over 30 nts

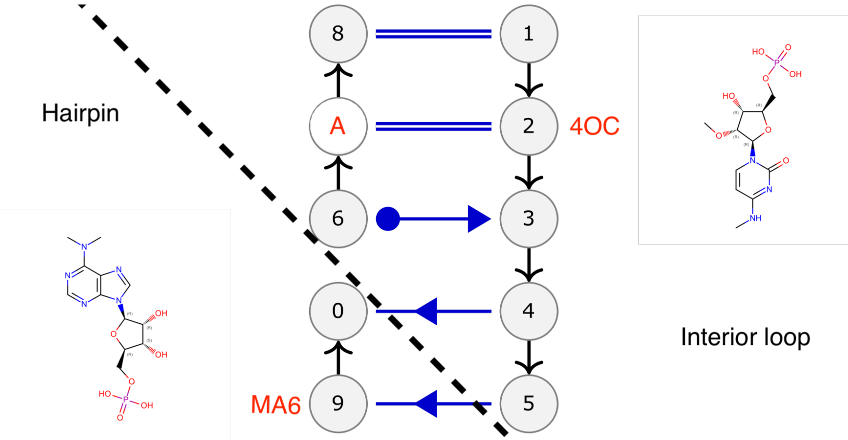


Distribution of type of RNAs with chemical modifications known 3D structures


9627 modified nts
RNAs over 30 nts

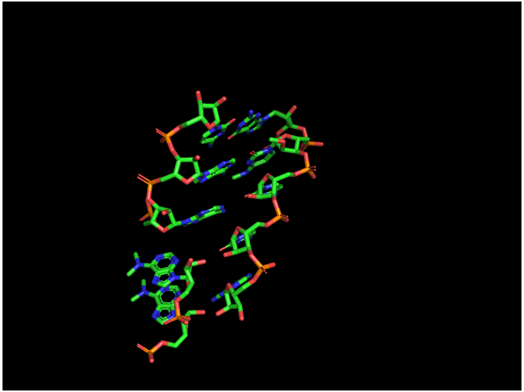
Chemical modification necessary for these modules

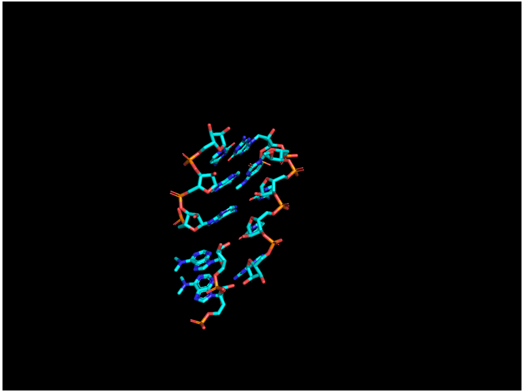
(Thanks Craig Zirbel)



- 7PWO — 18S G. Lamblia
- 4Y4O — 16S T. Thermophilus
- 7RYG — 16S A. baumannii
- 7KGB — 16S M. Tuberculosis
- 7UNW — 16S P. Aeruginosa
- 5J7L — 16S E. coli

G. Lamblia vs E. Coli

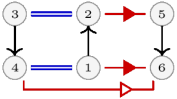




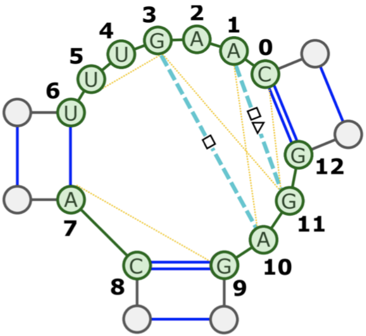
What can we do with that?

Identification in sequence

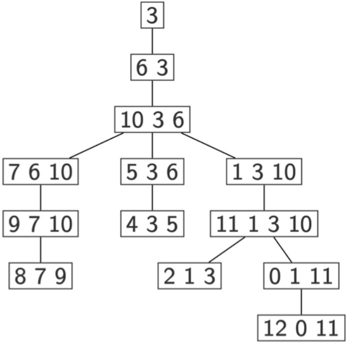
- If we know many sequences with same sub-structure
- Sometimes yes, sometimes no for loops (Sarrazin-Gendron et al., NAR, 2019, RECOMB 2020)
- Challenge with low sequence specificity (A-minor wants one side to be stack cWW)
- Can we leverage structural context? (Gianfrotta et al. SEA 2021)



BayesPairing Bayesian Network



(a) RNA module



(b) Tree decomposition

$$\mathbb{P}(p_3 = x), \forall x \in \mathcal{B} = \{\mathsf{A}, \mathsf{C}, \mathsf{G}, \mathsf{U}\}$$

$$\mathbb{P}(p_6 = x \mid p_3 = y), \forall (x, y) \in \mathcal{B}^2$$

$$\mathbb{P}(p_{10} = x \mid p_3, p_6 = y, y')$$

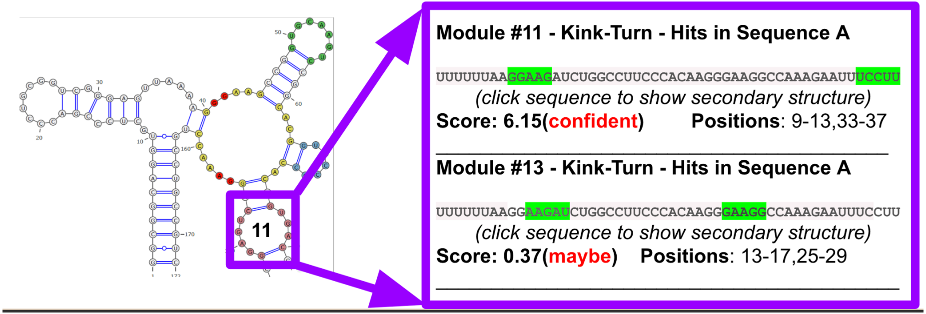
$$\mathbb{P}(p_7 = x \mid p_6, p_{10} = y, y')$$

$$\vdots$$

$$\mathbb{P}(p_{11} = x \mid p_1, p_3, p_{10} = y, y', y'')$$

(c) Conditional probabilities

Combine with structure prediction



Module #11 - Kink-Turn - Hits in Sequence A

UUUUUUAAGGAAGAUCUGGCCUUCCCACAAGGGAAGGCCAAAGAAUUUCCUU

(click sequence to show secondary structure)

Score: 6.15(confident) **Positions**: 9-13,33-37

Module #13 - Kink-Turn - Hits in Sequence A

UUUUUUAAGGAAGAUCUGGCCUUCCCACAAGGGAAGGCCAAAGAAUUUCCUU

(click sequence to show secondary structure)

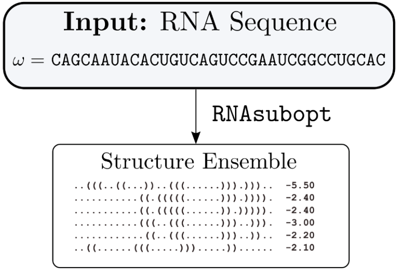
Score: 0.37(maybe) **Positions**: 13-17,25-29

BayesPairing workflow

Input: RNA Sequence

$\omega =$ `CAGCAAUACACUGUCAGUCCGAAUCGGCCUGCAC`

BayesPairing workflow



Input: RNA Sequence

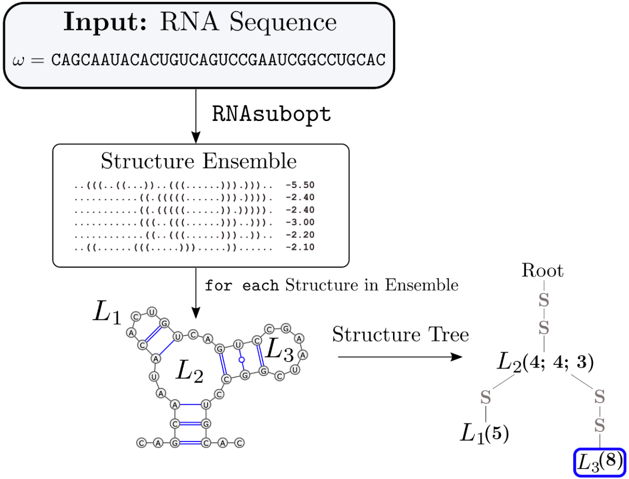
$\omega =$ `CAGCAAUACACUGUCAGUCCGAAUCGGCCUGCAC`

`RNAsubopt`

Structure Ensemble

```
..(((..((...))..(((......))).)))..   -5.50
...........((.(((((......))).)))).   -2.40
...........((.(((((......)).))))).   -2.40
...........(((..(((......)))..))).   -3.00
............((..(((......)))..))..   -2.20
..((......(((.....))).....))......   -2.10
```

BayesPairing workflow



Input: RNA Sequence

$\omega =$ `CAGCAAUACACUGUCAGUCCGAAUCGGCCUGCAC`

`RNAsubopt`

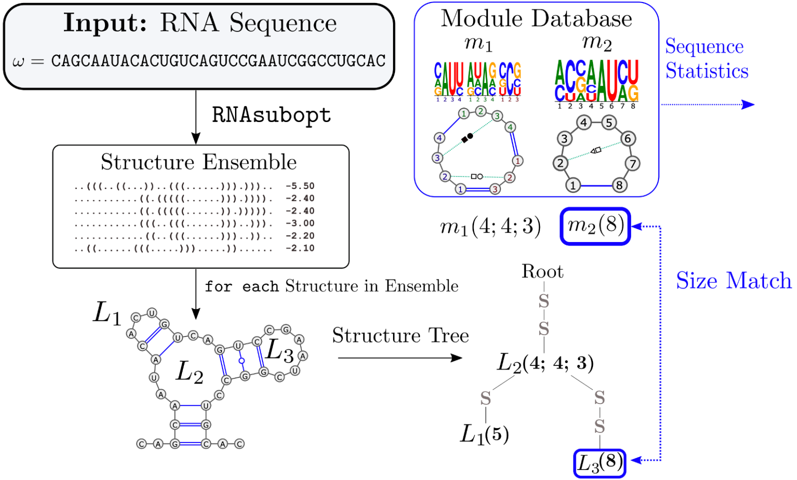
Structure Ensemble

```
..(((..((...))..(((......))).)))..  -5.50
...........((.(((((......))).)))).  -2.40
...........((.(((((......)).))))).  -2.40
...........(((..(((......)))..))).  -3.00
............((..(((......)))..))..  -2.20
..((......(((.....))).....))......  -2.10
```

`for each` Structure in Ensemble

Structure Tree

BayesPairing workflow



Input: RNA Sequence

$\omega =$ `CAGCAAUACACUGUCAGUCCGAAUCGGCCUGCAC`

`RNAsubopt`

Structure Ensemble

```
..(((..((...))..(((......))).)))..  -5.50
...........((.(((((......))).)))).  -2.40
...........((.(((((......)).))))).  -2.40
...........(((..(((......)))..))).  -3.00
............((..(((......)))..))..  -2.20
..((......(((.....))).....))......  -2.10
```

`for each` Structure in Ensemble

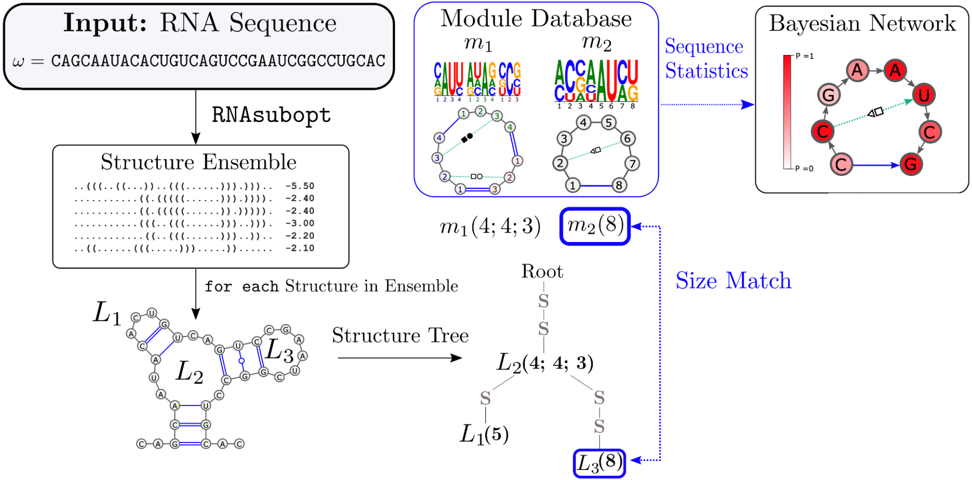
Structure Tree

Module Database

Sequence Statistics

Size Match

BayesPairing workflow



Input: RNA Sequence

$\omega =$ `CAGCAAUACACUGUCAGUCCGAAUCGGCCUGCAC`

`RNAsubopt`

Structure Ensemble

```
..(((..((...))..(((......))).)))..   -5.50
...........((.(((((......))).)))).   -2.40
...........((.(((((......)).))))).   -2.40
...........(((..(((......)))..))).   -3.00
............((..(((......)))..))..   -2.20
..((......(((.....))).....))......   -2.10
```

`for each` Structure in Ensemble

L_1 L_2 L_3

Structure Tree

Root

$L_2(4;\,4;\,3)$

$L_1(5)$

$L_3(8)$

Module Database

m_1 m_2

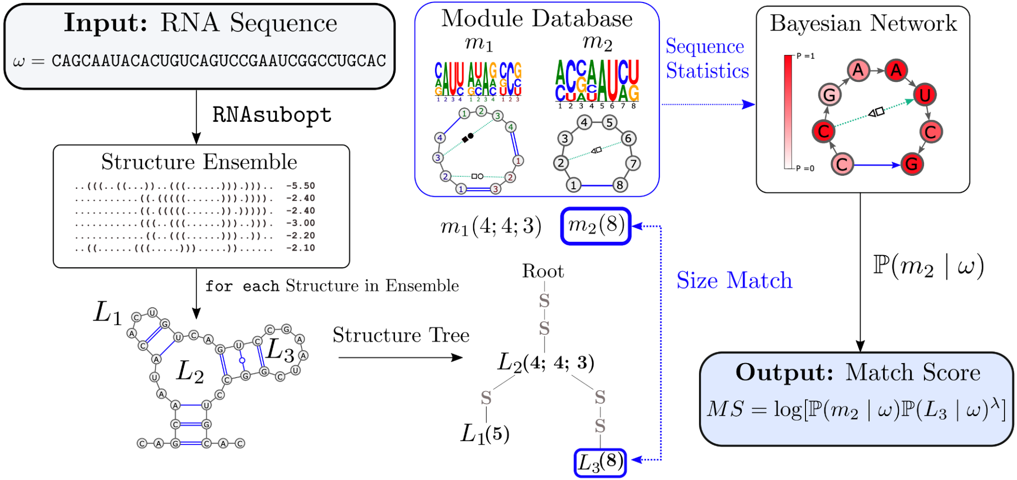
$m_1(4;4;3)$ $m_2(8)$

Sequence Statistics

Size Match

Bayesian Network

BayesPairing workflow



Input: RNA Sequence

$\omega =$ `CAGCAAUACACUGUCAGUCCGAAUCGGCCUGCAC`

`RNAsubopt`

Structure Ensemble

```
..(((..((...))..(((......))).)))..   -5.50
...........((.(((((......))).)))).   -2.40
...........((.(((((......)).))))).   -2.40
...........(((..(((......)))..))).   -3.00
............((..(((......)))..))..   -2.20
..((......(((.....))).....))......   -2.10
```

`for each` Structure in Ensemble

Structure Tree

Module Database

Sequence Statistics

Bayesian Network

Size Match

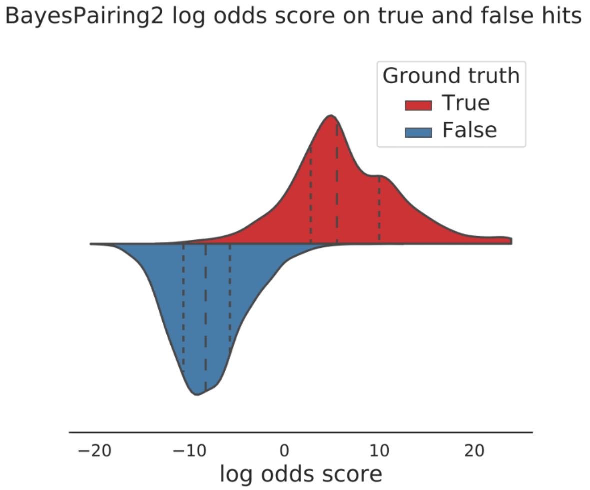
$\mathbb{P}(m_2 \mid \omega)$

$m_1(4;4;3)$ $\boxed{m_2(8)}$

Output: Match Score

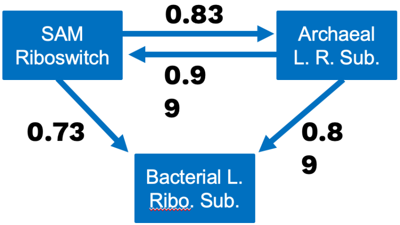
$$MS = \log[\mathbb{P}(m_2 \mid \omega)\mathbb{P}(L_3 \mid \omega)^\lambda]$$

BayesPairing is discriminant



BayesPairing learns across families

Kink-turn



SAM Riboswitch

Archaeal L. R. Sub.

Bacterial L. Ribo. Sub.

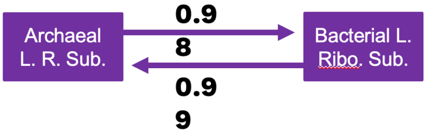
0.83

0.99

0.73

0.89

Sarcin-Ricin



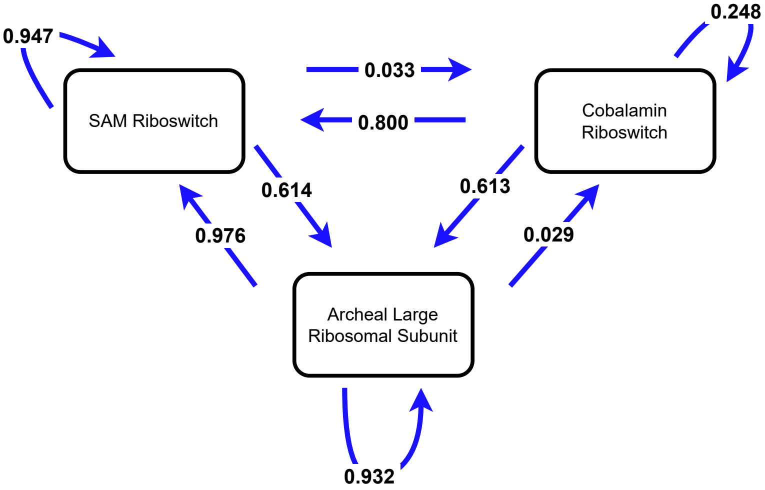
Archaeal L. R. Sub.

Bacterial L. Ribo. Sub.

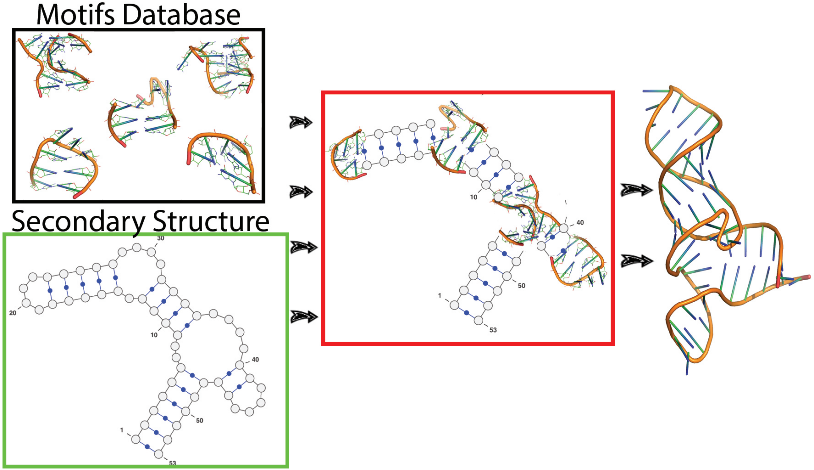
0.98

0.99

BayesPairing with KinkTurn (Thanks Anton Petrov and Nancy Ontiveros)



Towards full 3D structure reconstruction

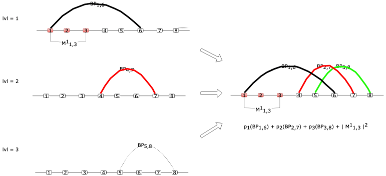


Motifs Database

Secondary Structure

And now?

- Extend BayesPairing for long-range motifs (R. Sarrazin-Gendron is doing it now)
- Find motifs variations
- Extend integer programming framework for concurrent prediction of pseudoknotted-structure and complex motifs (G. Loyer is doing it now)



- What is the flexibility of motifs? And how does it impact fonction?

Acknowledgment

- Antoine Soulé
- Jérôme Waldipühl
- Tsvi Tlusty
- Roman Sarrazin-Gendron

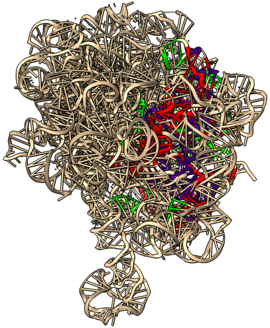
- Alain Denise
- Eric Westhof
- François Major
- Gabriel Loyer

NSERC
CRSNG

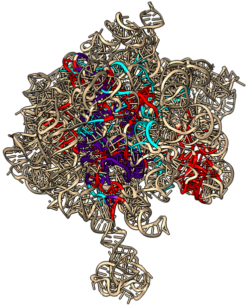
Fonds de recherche
Santé
Québec

FSM

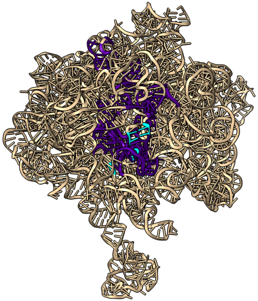
And also thanks to:
Craig Zirbel, Nancy Otiveros, Yann Ponty, Anton Petrov, Leontis Neocles



T. Thermophilus 70S

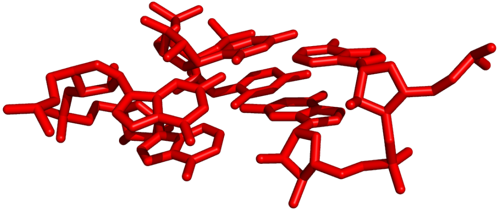


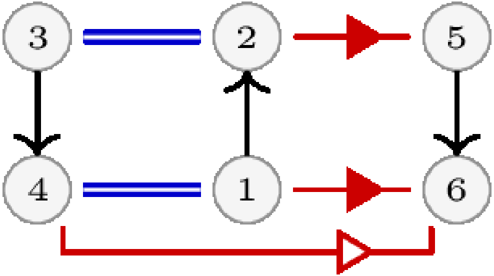
E. Coli 70S



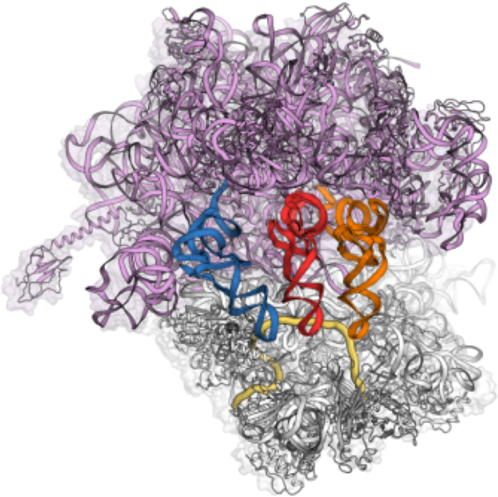
P. Aeruginosa 50S

A-minor module



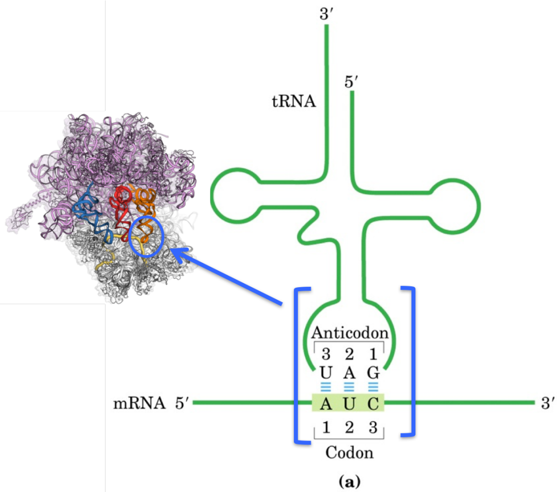


Multi-chains
Ribosomal translation



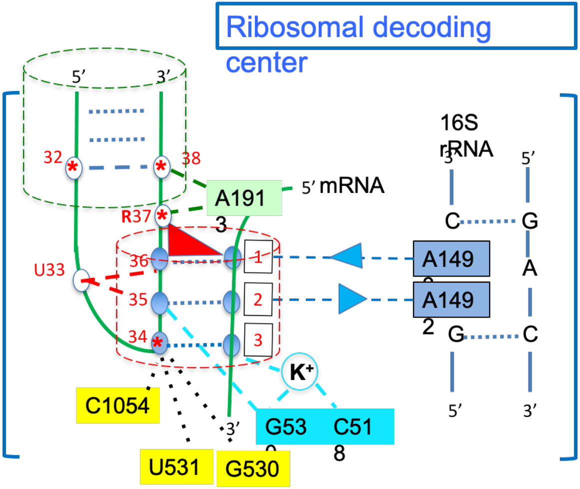
from E. Westhof

Multi-chains
Ribosomal translation



from E. Westhof

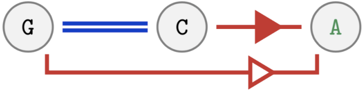
Multi-chains
Ribosomal translation

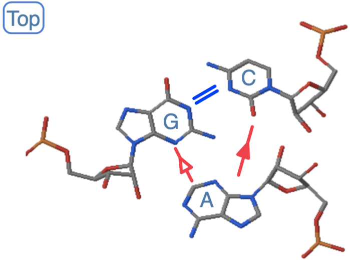


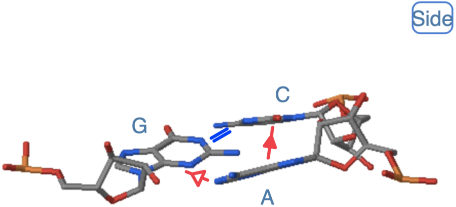
Ribosomal decoding center

from E. Westhof

Triplet







C vs 4OC (O2’-Methylcytidine-5’-monophosphate)