

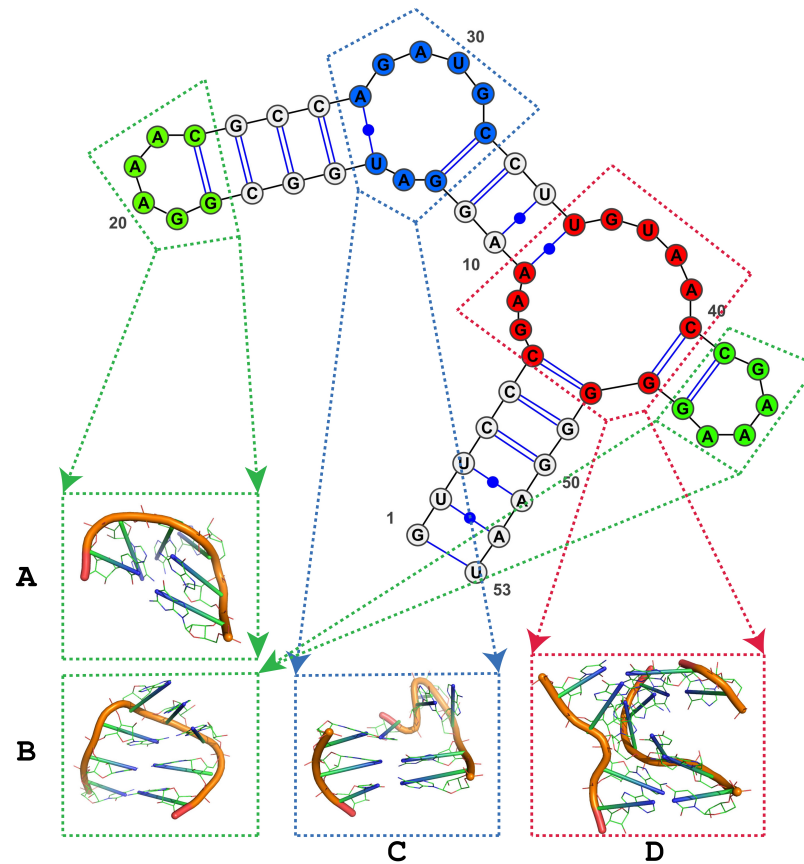
Nucleotide-nucleotide interaction networks

Vladimir Reinhartz

Université du Québec à Montréal



Loops are organized



Non-canonical base pairs

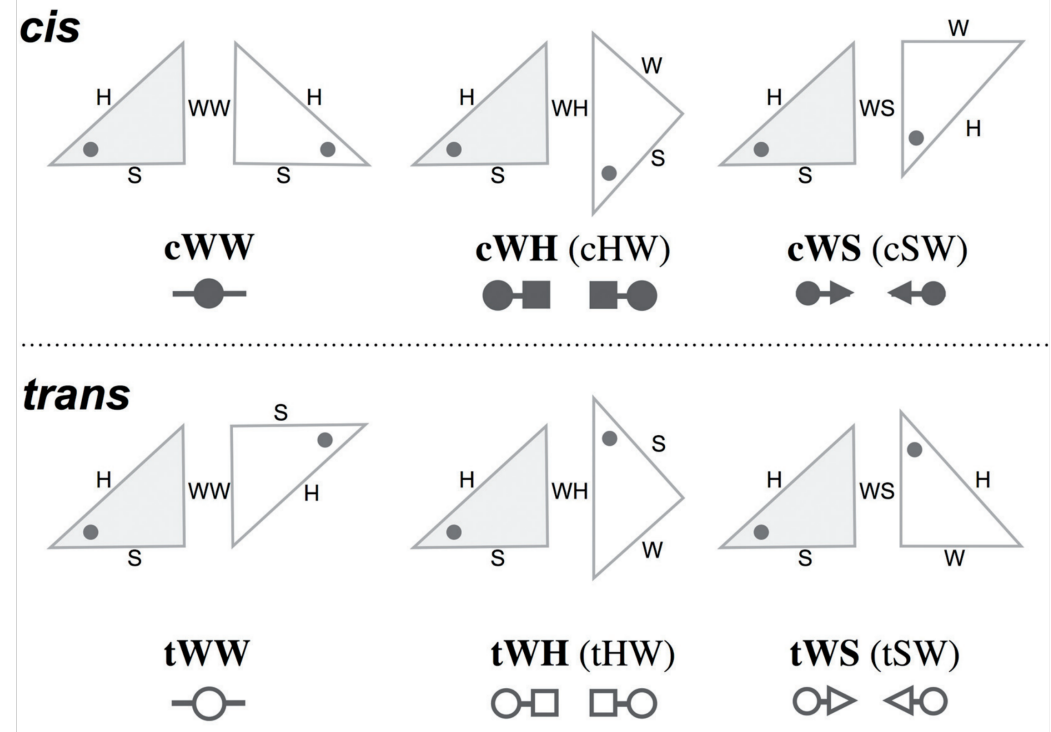
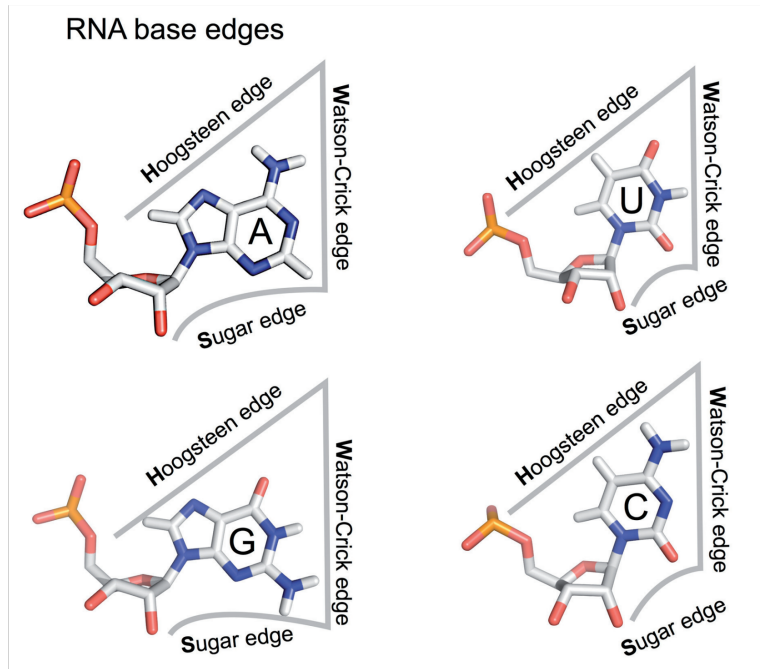
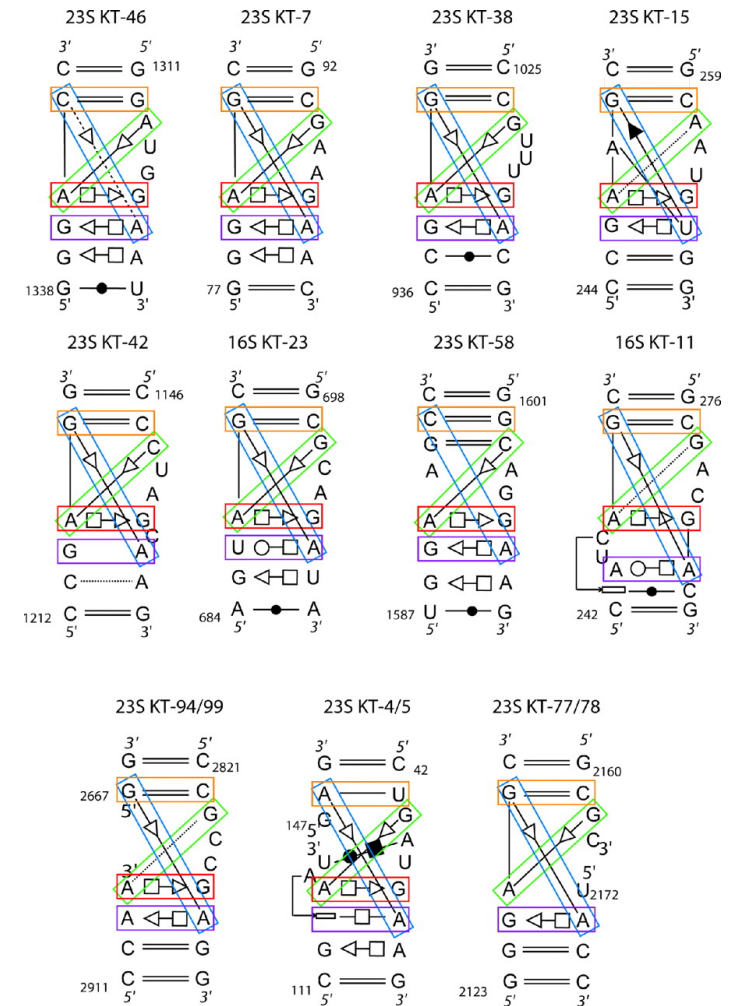
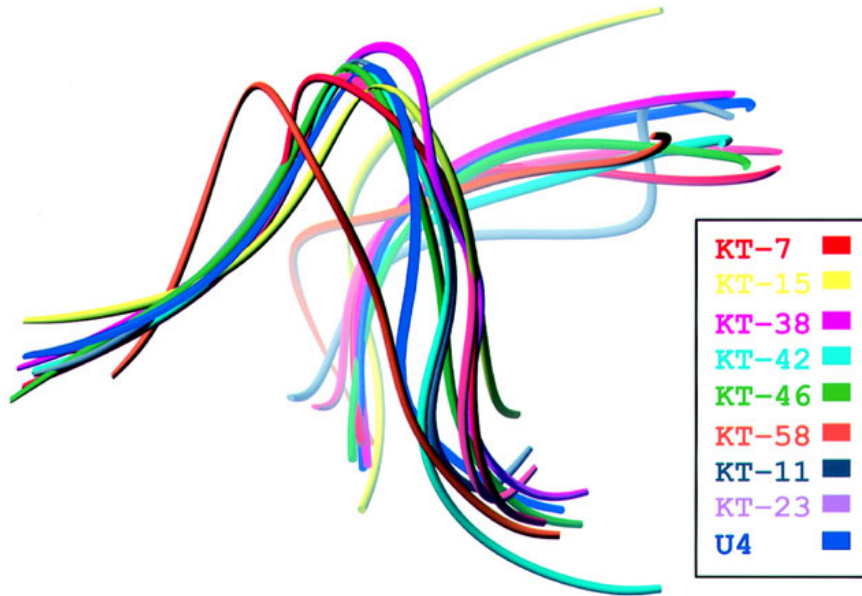


Figure adapted from
Almakarem et al, 2011

Conserved basepairs interaction networks



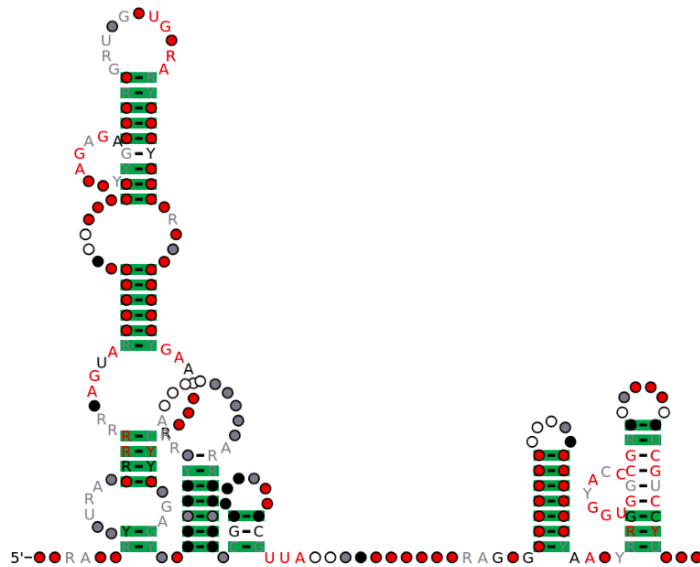
RNA modules

No sequence only structure

RNA motifs modules are directed and ordered stacked arrays of non-Watson–Crick base pairs forming distinctive foldings of the phosphodiester backbones of the interacting RNA strands

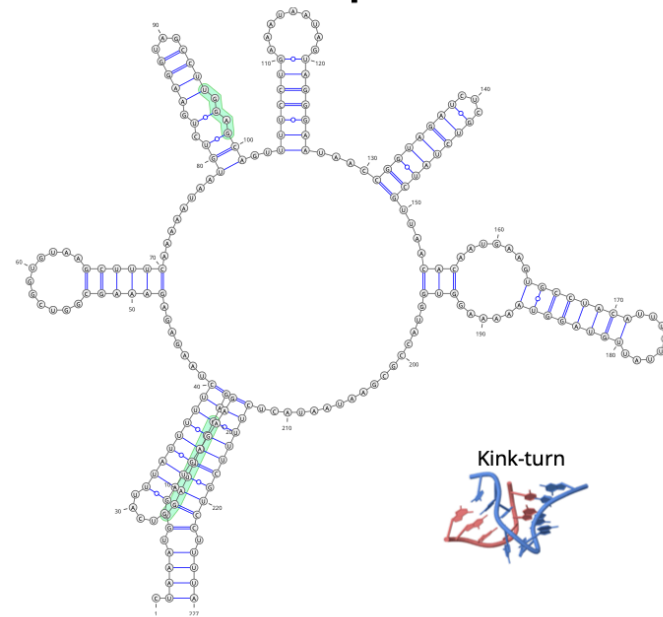
Why modules matter

T-box Consensus structure (RF00230)



(Taken from Rfam)

T-box MFE prediction

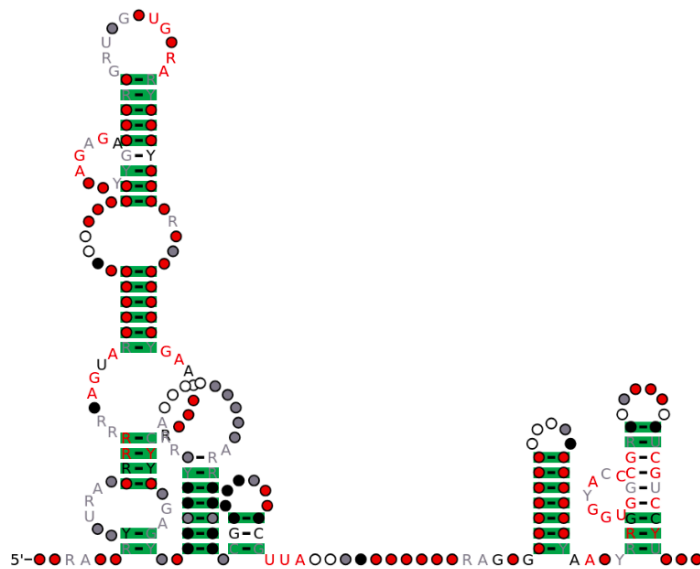


Kink-turn



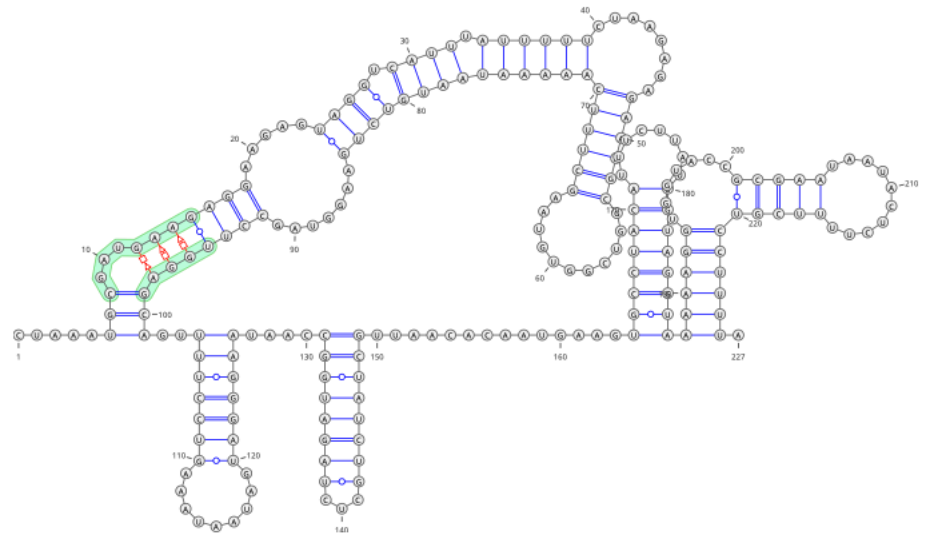
Why modules matter

T-box Consensus structure (RF00230)



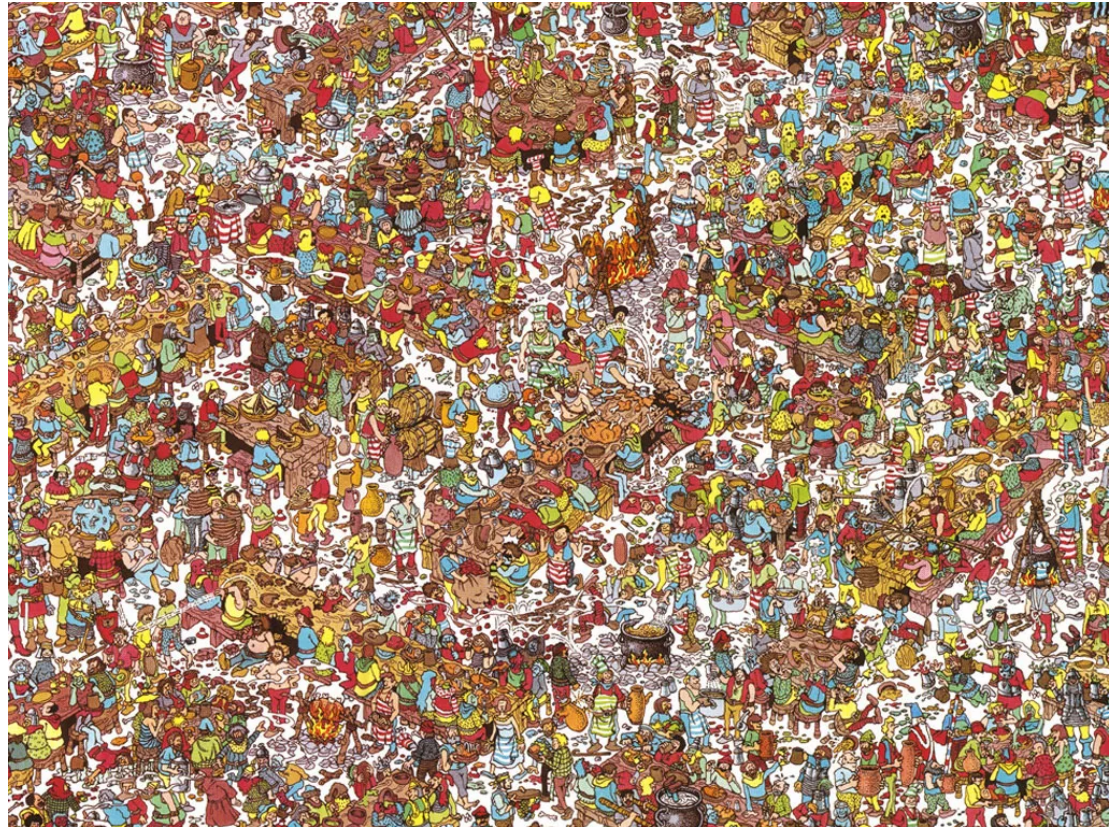
(Taken from Rfam)

MFE + hard constraint



What are the modules?

Where's Waldo? (Or Wally or Walter or Charlie or Holger or Уолли?)



What are the modules?

Inverse where's Waldo? (Or Wally or Walter or Charlie or Holger or Уолли ?) !
From two images, what is in common?

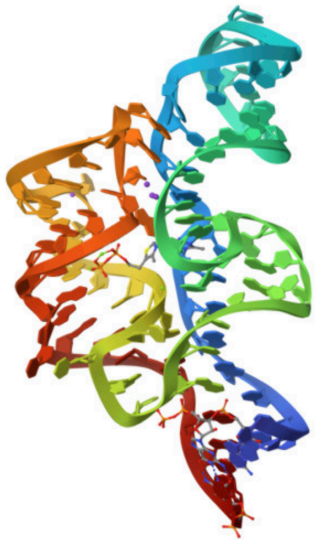


RNA as graphs

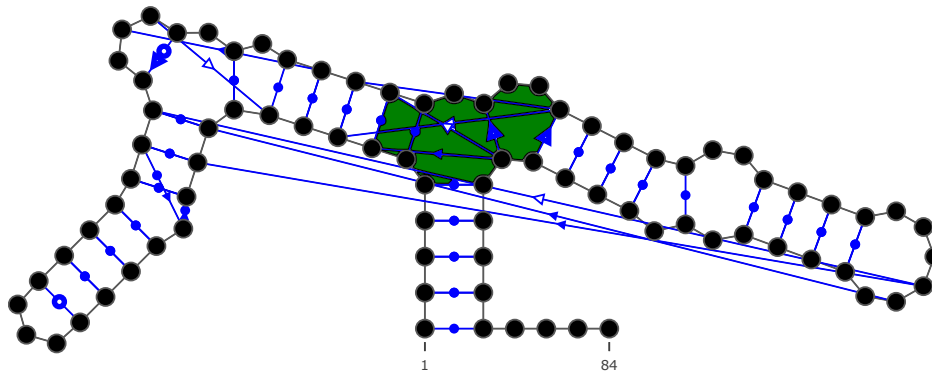
*only phosphodiester bonds and interactions

**We also have stackings but there is too much data (>30TB in DB)

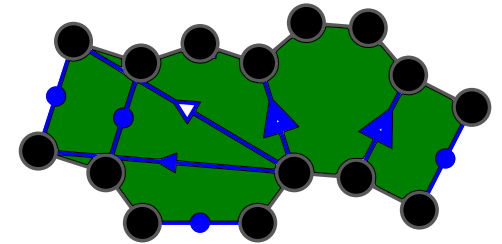
Based on non-redundant dataset 4.22
at 3Å from RNA3DHub on 2026/01/14
which has 2 405 structures



2GDI



Global structure

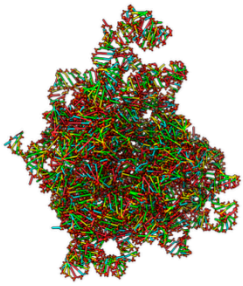


Loop
(3-way junction)

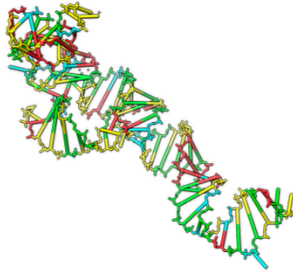
Large variety of RNA

We keep chain-chain interactions

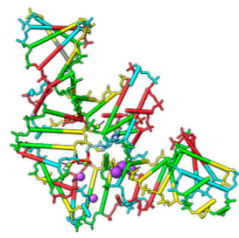
FR3D annotations (for now?)



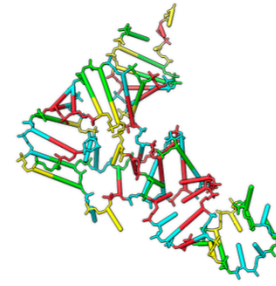
23S rRNA



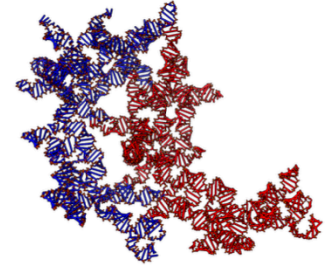
5S rRNA



TPP riboswitch

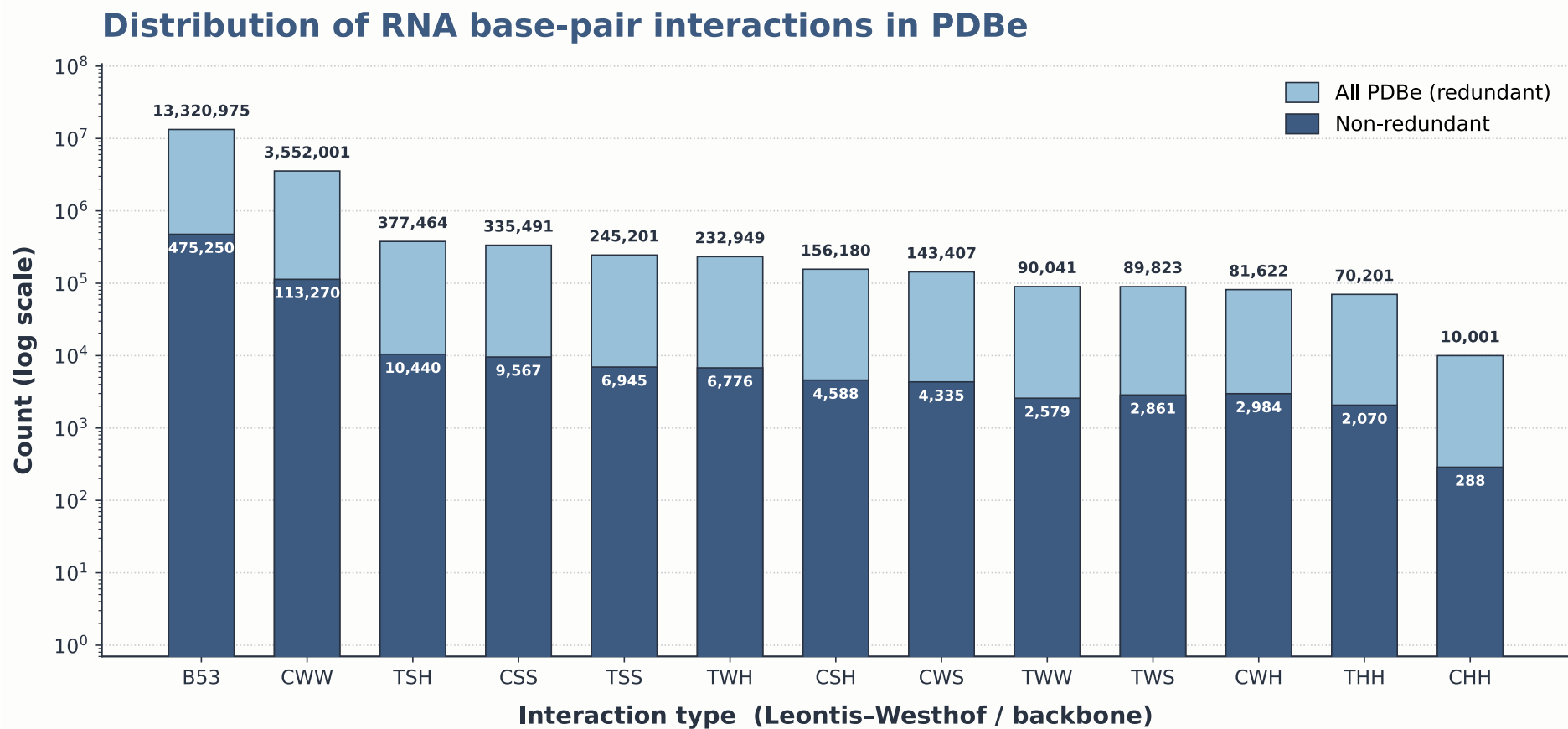


tRNA



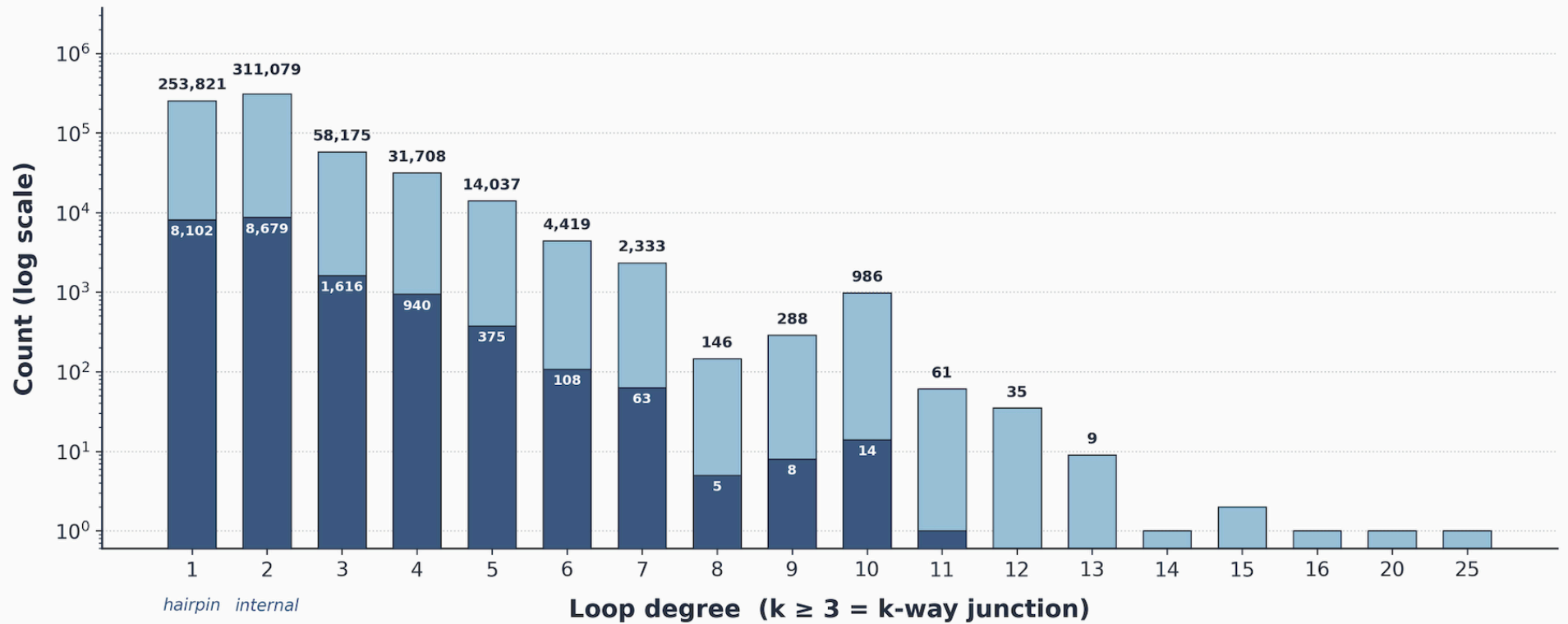
GOLLD lncRNA

Interactions diversity



Loops diversity

B. Loop degrees in RNA structures



Given arbitrary graphs, what are all conserved subgraphs? (enumeration of maximal subgraph isomorphism)

- NP-Hard
- Can be exponential in nodes

But

- Biological structures have low degree
- We have lots of computers

Given arbitrary graphs, what are all conserved subgraphs? (enumeration of maximal subgraph isomorphism)

- NP-Hard
- Can be exponential in nodes

But

- Biological structures have low degree
- We have lots of computers



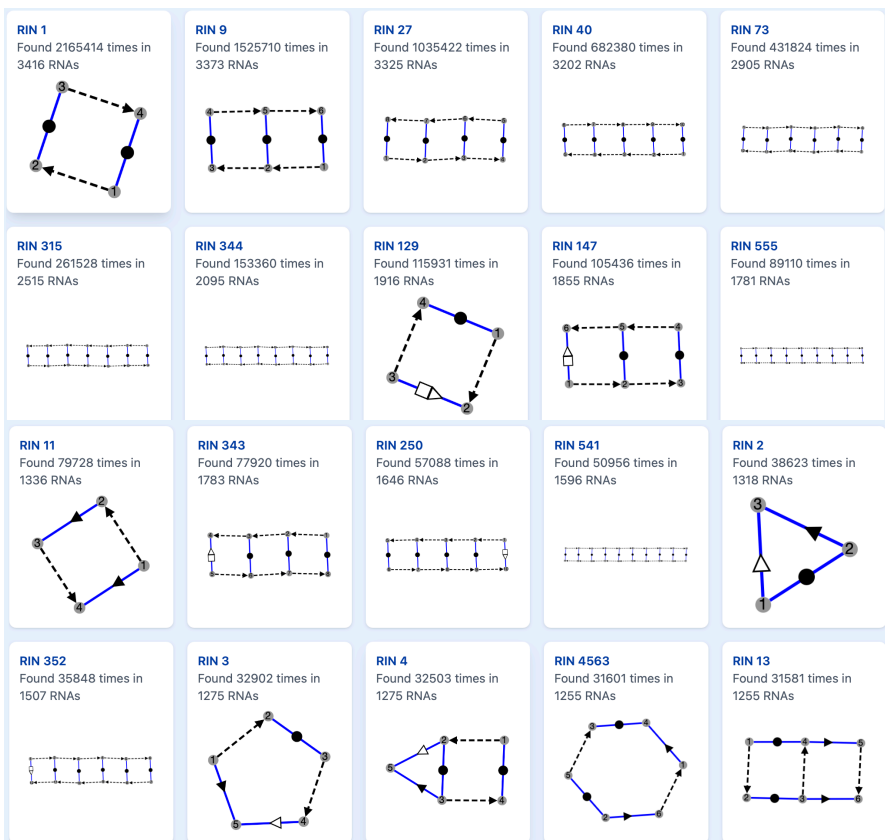
Wilfried Agbeto

Distributed system
and memory subgraph isomorphism

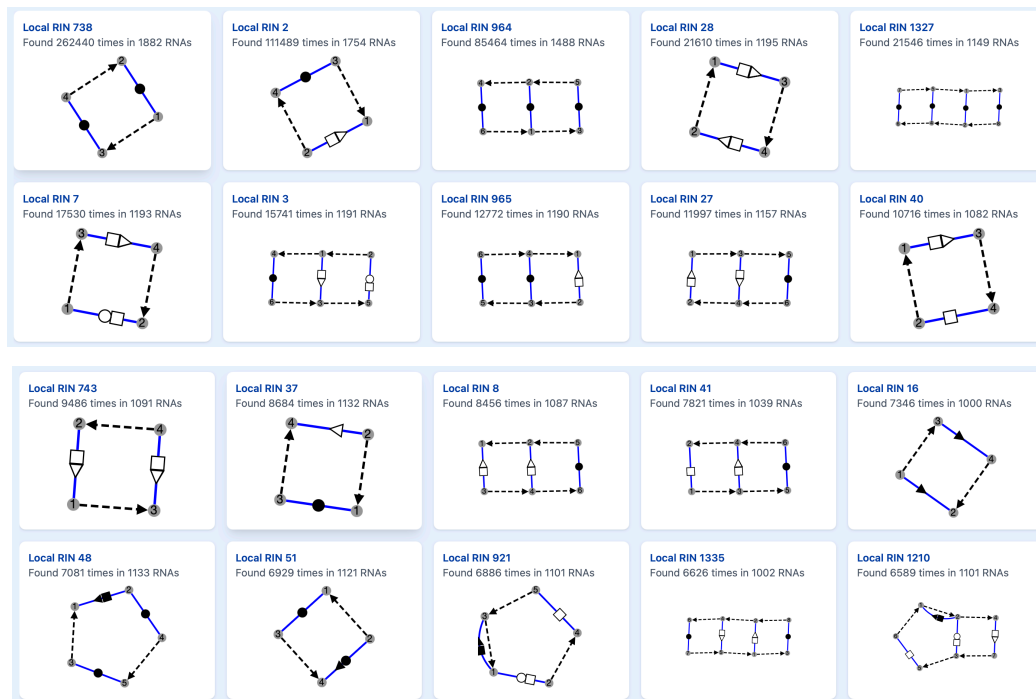
ComputeCanada cluster ->
3200 processors on 70 nodes
each with 128GB RAM for 4 weeks

Recurrent Interaction Network (RIN)

Global

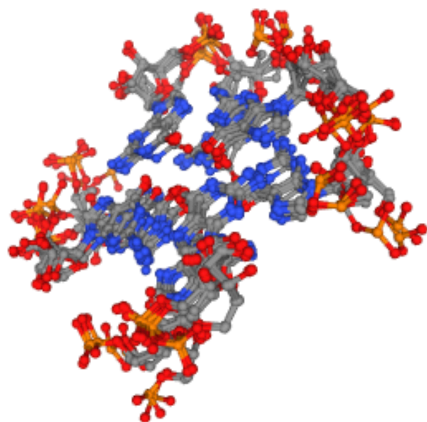
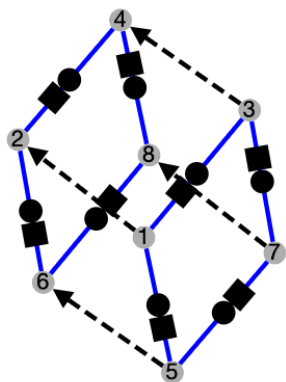


Local (loop)

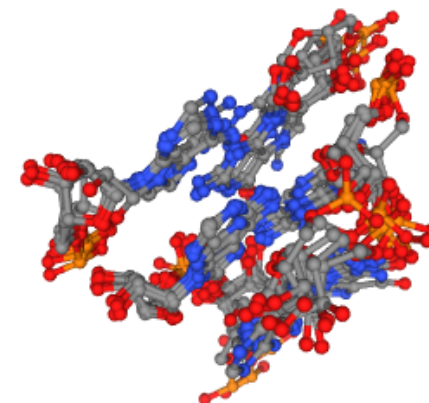
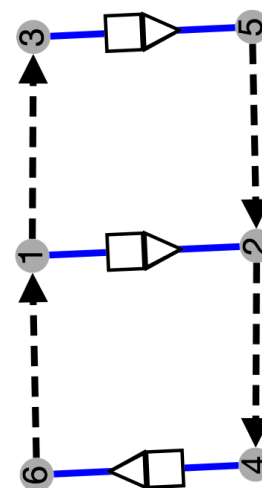


In structures under 3Å here

RIN 2438 - (mostly) G-quadruplex Observed 328 times



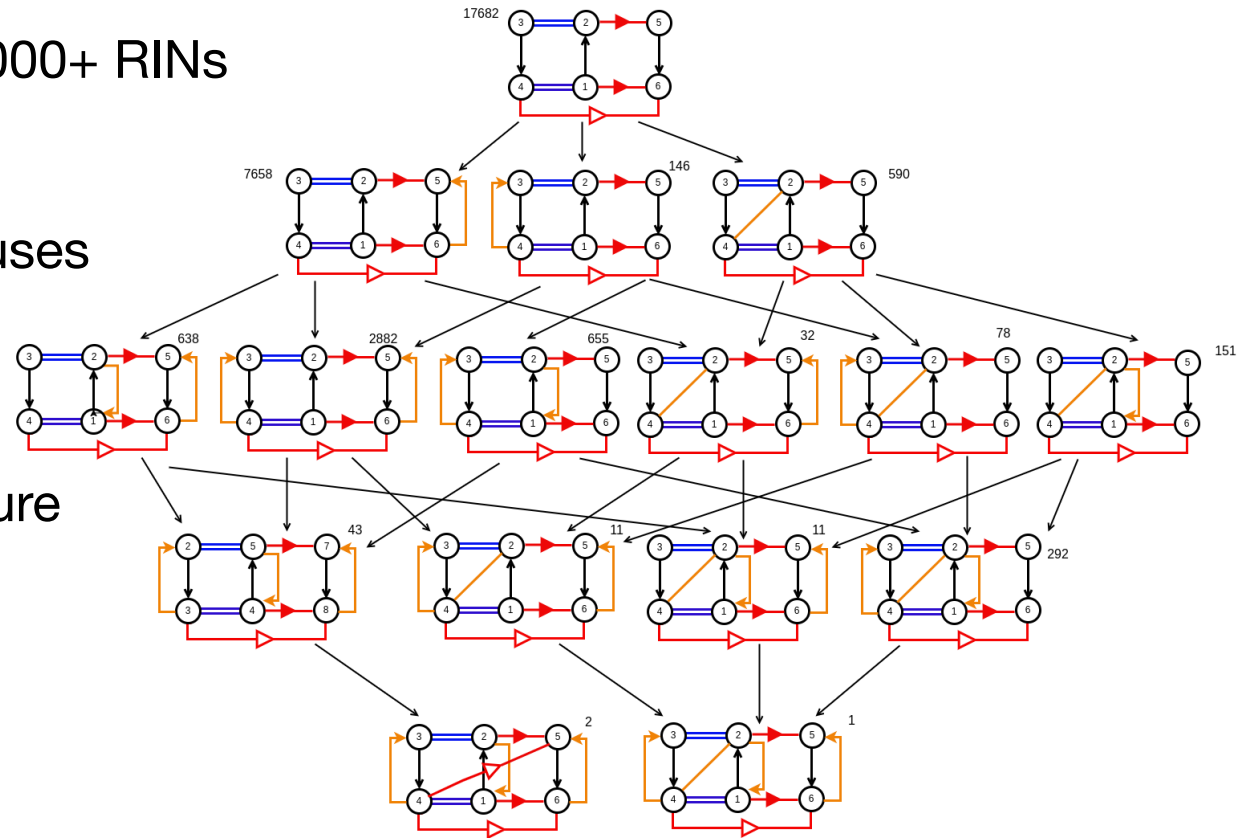
Local RIN 977 - kink-turn pattern Observed 4944 times



What about stacks?

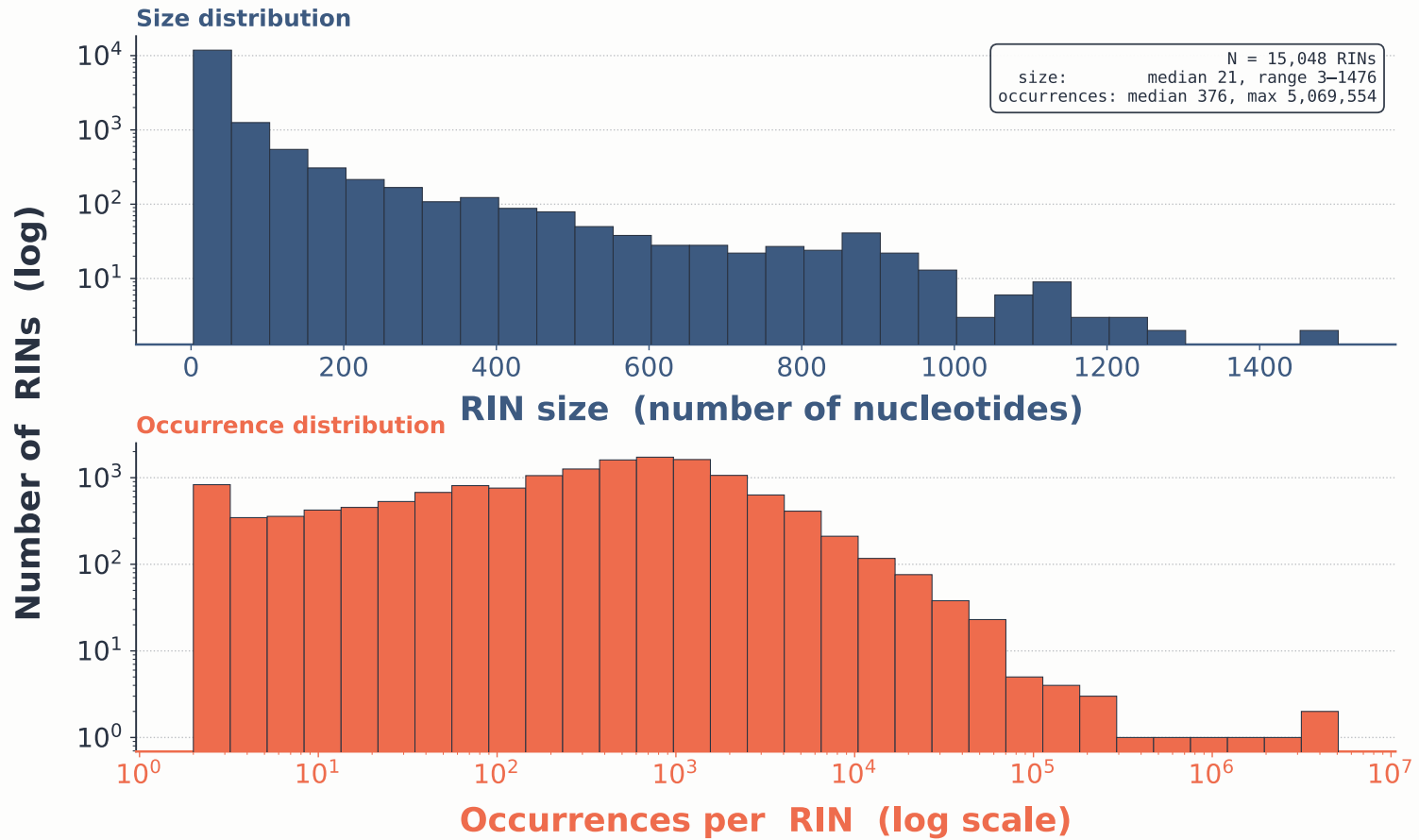
There's a lot

- Including stacks we jump to 150 000+ RINs (one order of magnitude)
- Annotating all there occurrences uses around 30TB of space
- For now not manageable, but they remain an important feature

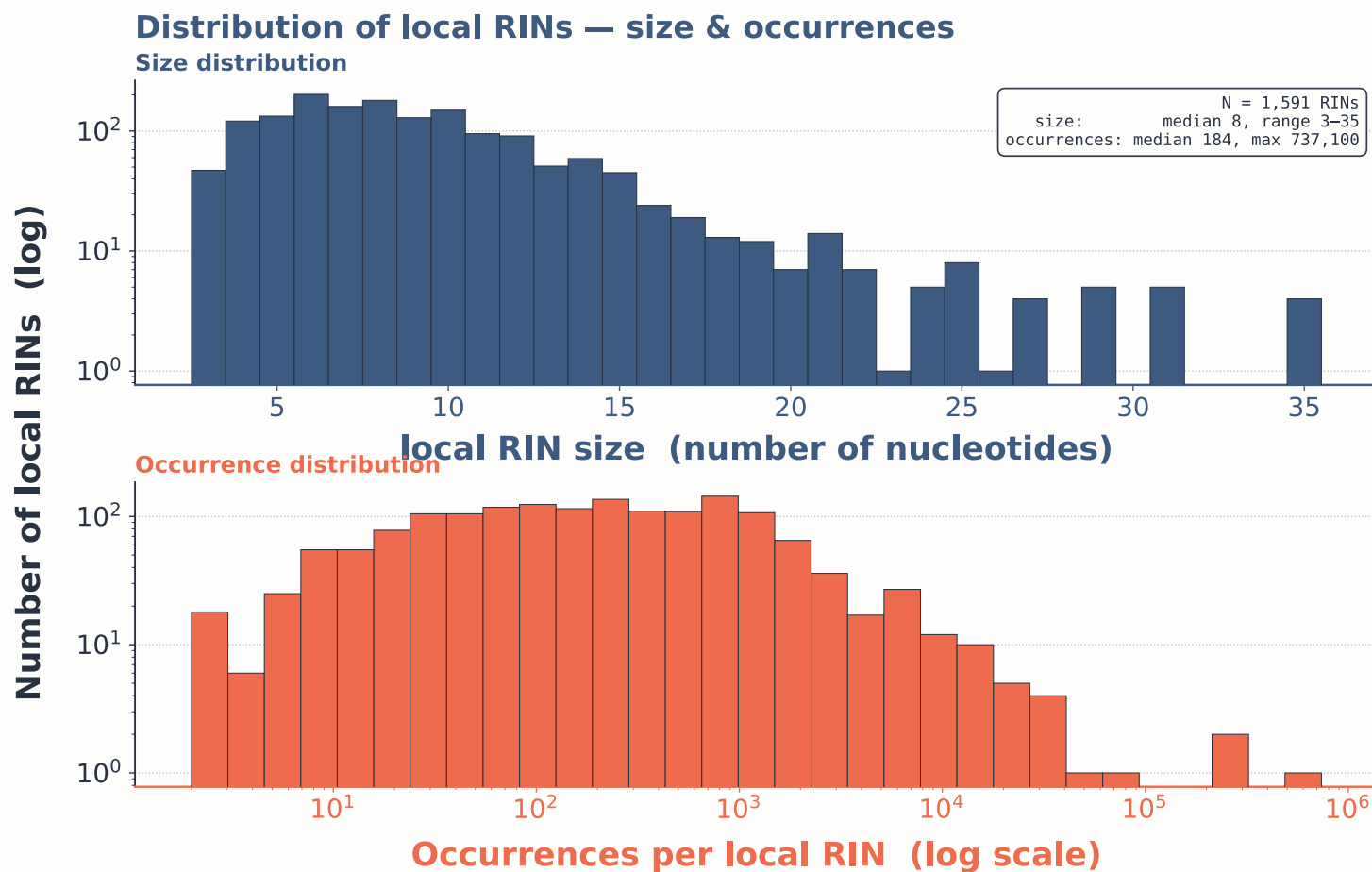


RIN global

Distribution of global RINs — size & occurrences



RIN local

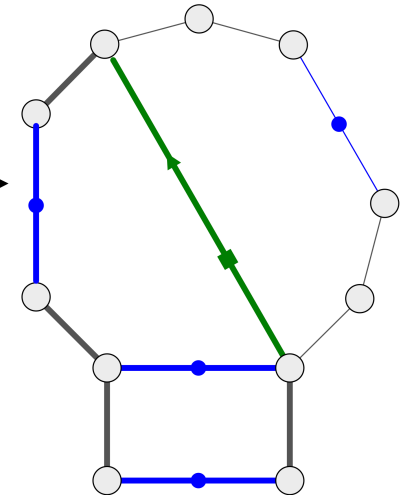
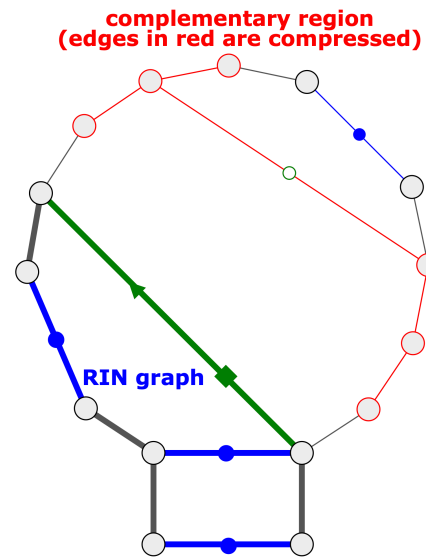
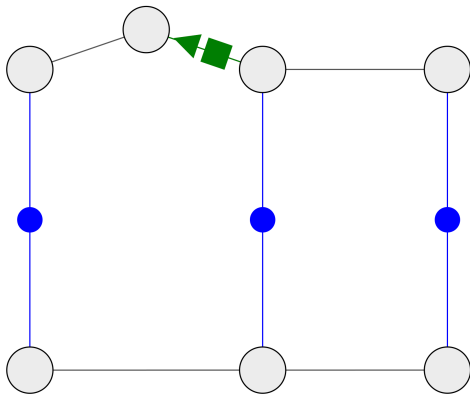


**For prediction, many algorithms work with
loops**

**How can we bridge the gap with
these recurrent interaction networks?**

Bulge Agnostic Local Loop Extended RIN Annotation BalleRINa

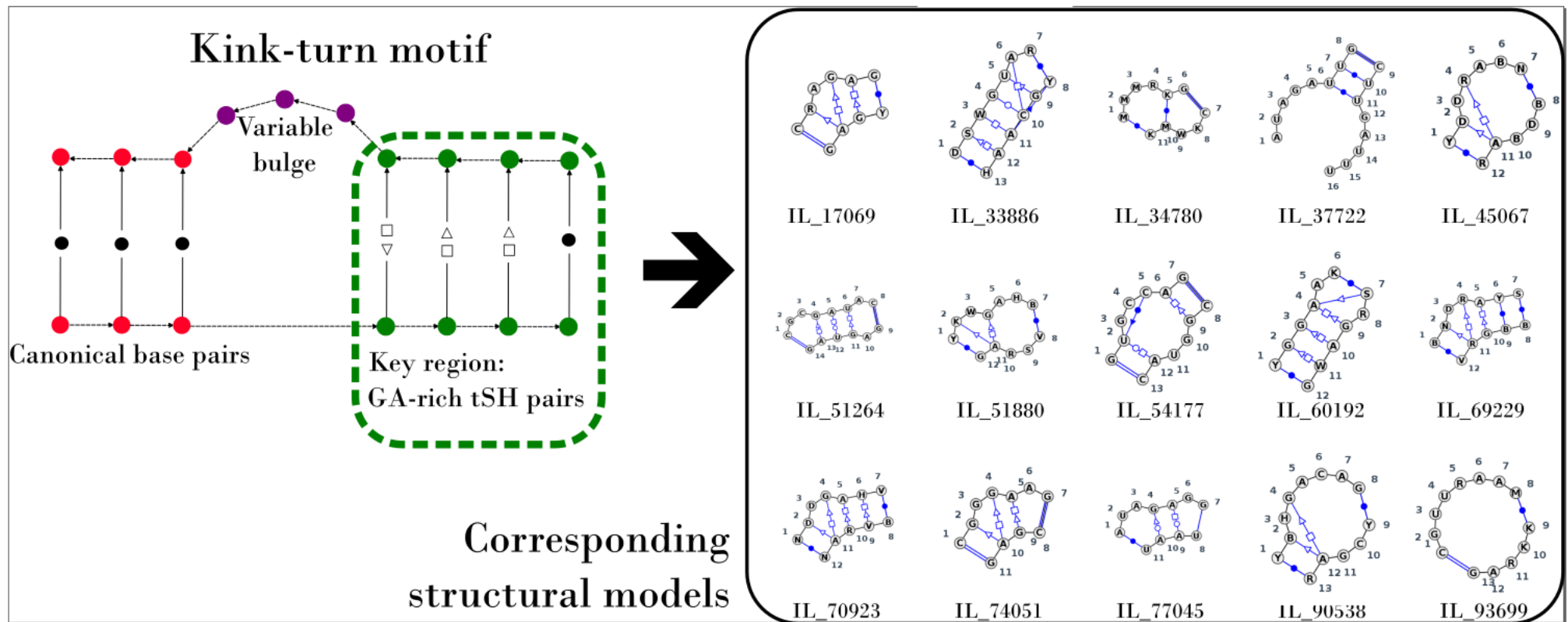
Recurrent Interaction Network



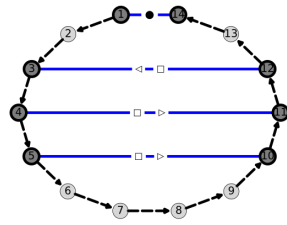
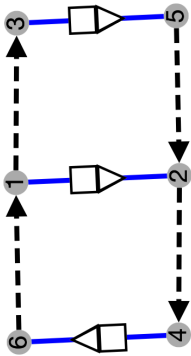
**RNA structure loop that
includes the RIN**

Simplified model of the loop

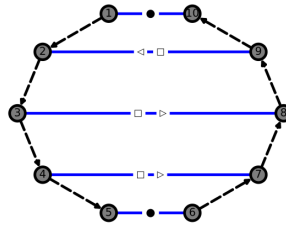
The kink-turn, an essential and common but very heterogeneous motif



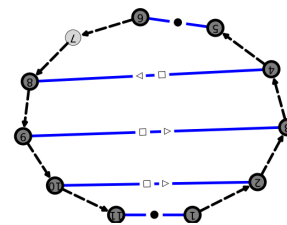
Kink-turn associated module -> many BalleRINas



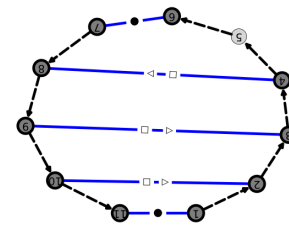
Hairpin



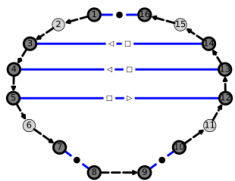
Interior Loop



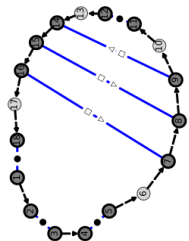
Interior Loop



Interior Loop



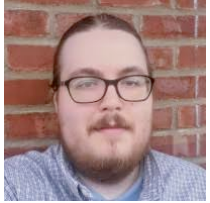
3-way junction



4-way junction

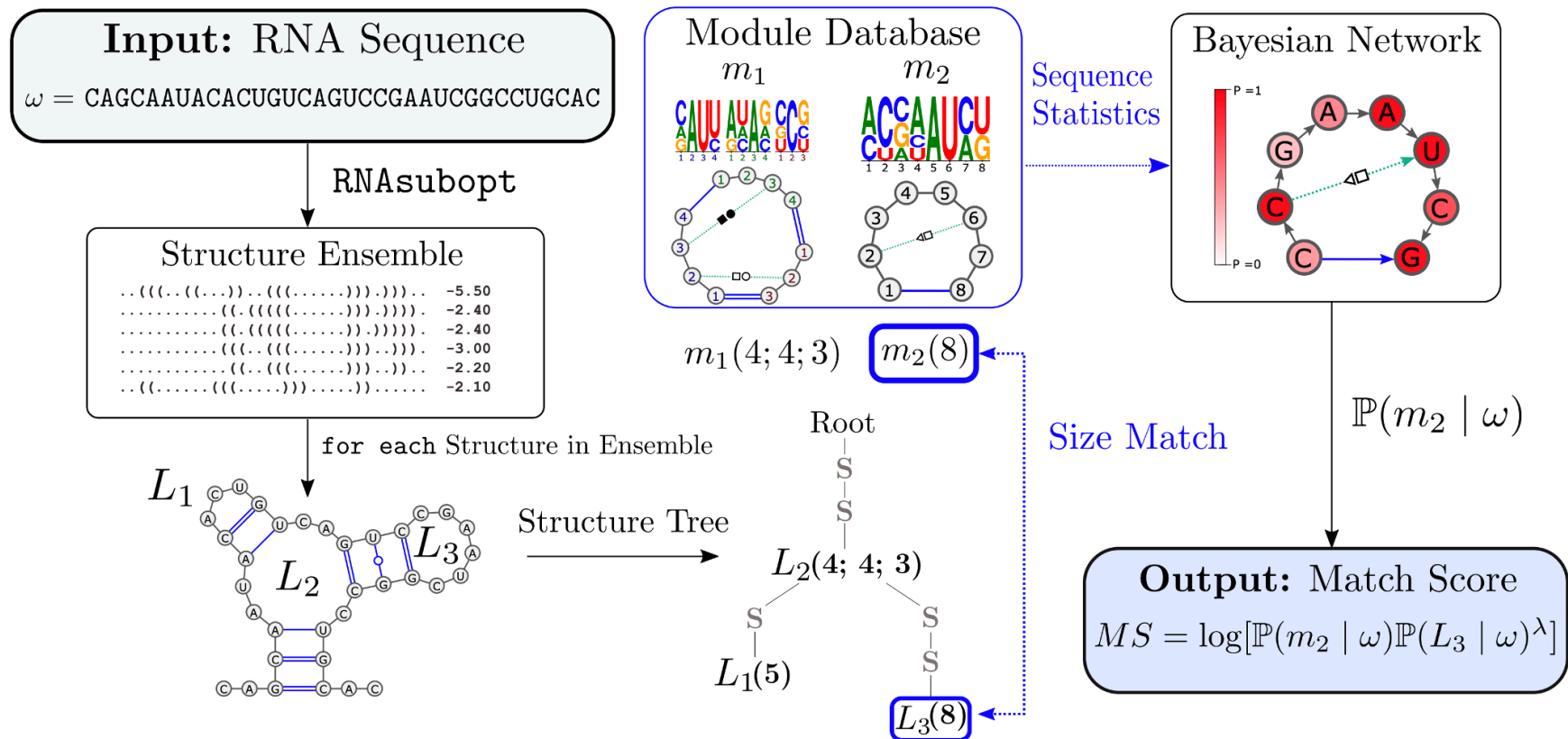


Hua-Ting Yao

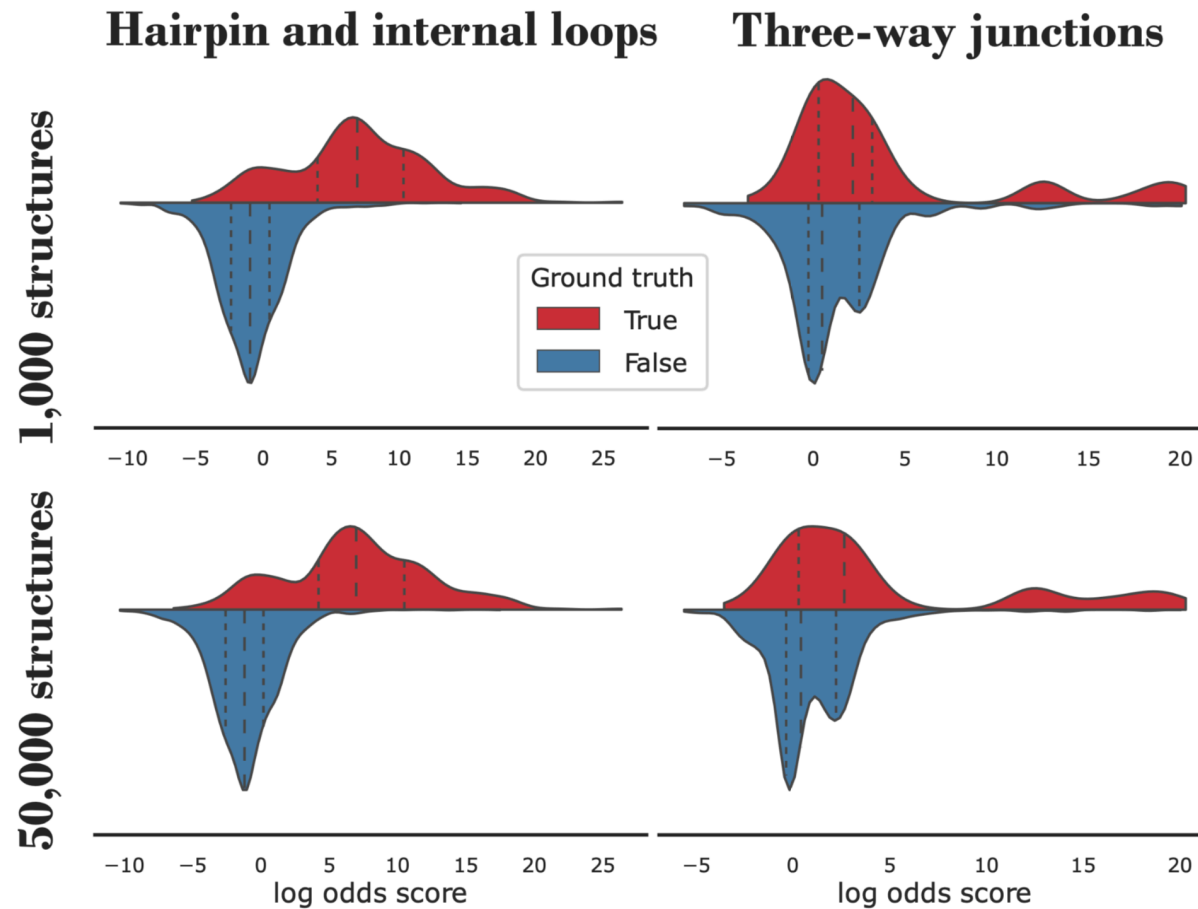


Roman Sarrazin-Gendron

Scoring loops with Bayesian Networks in a single sequence

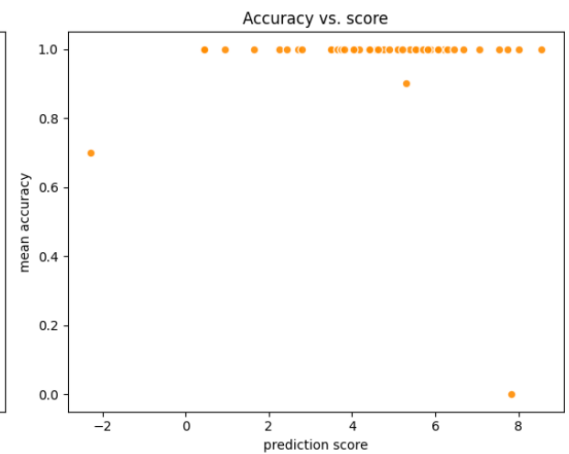
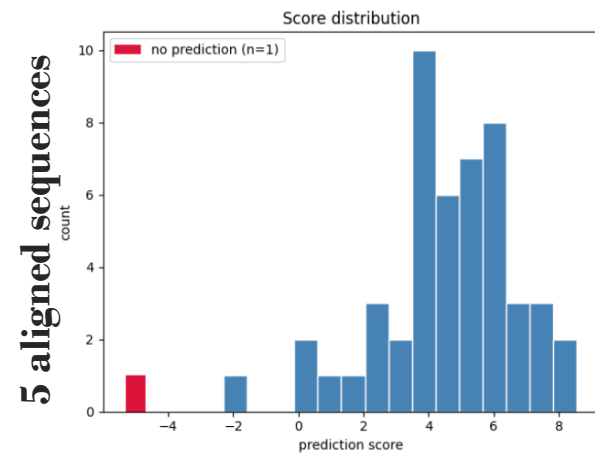
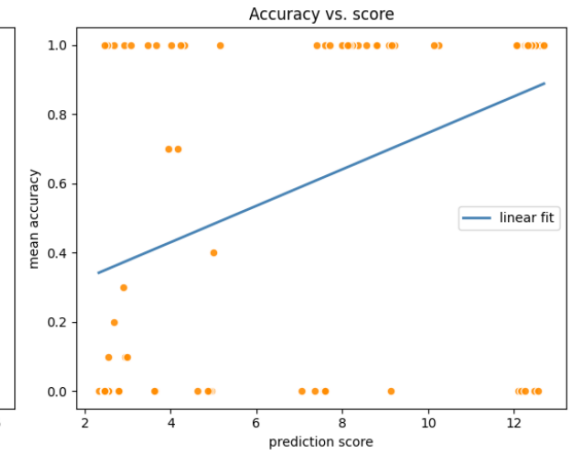
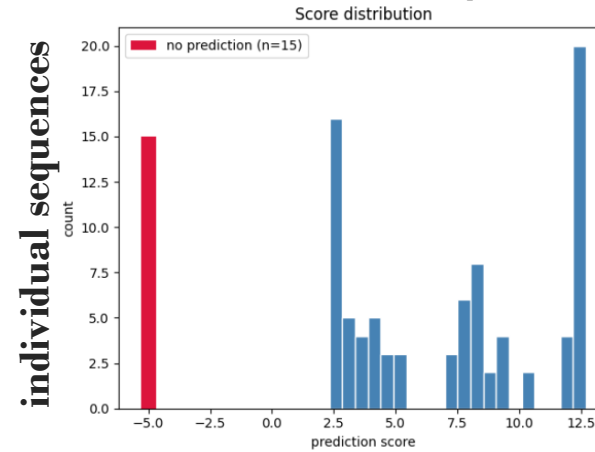
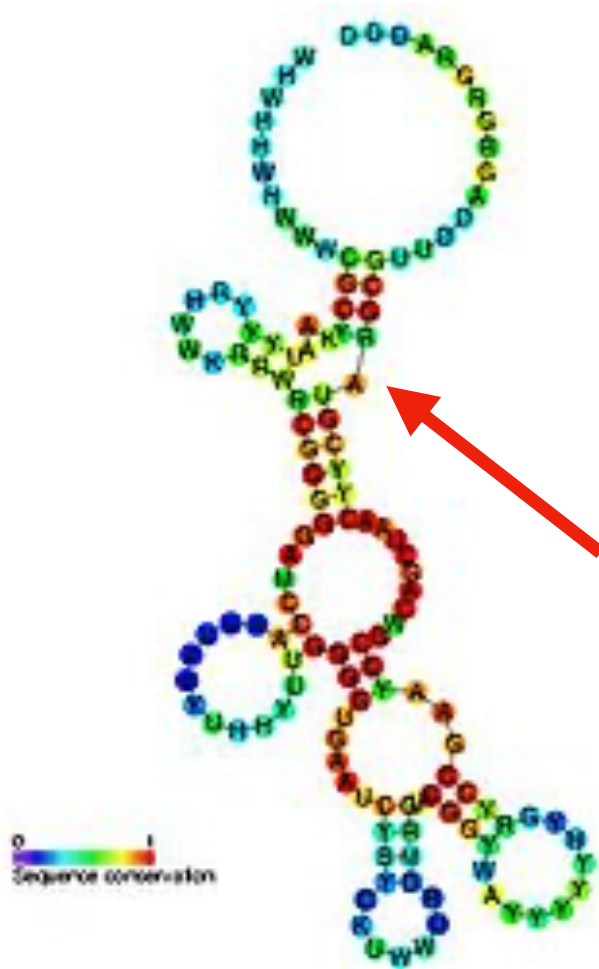


BayesPairing returns a log odds score



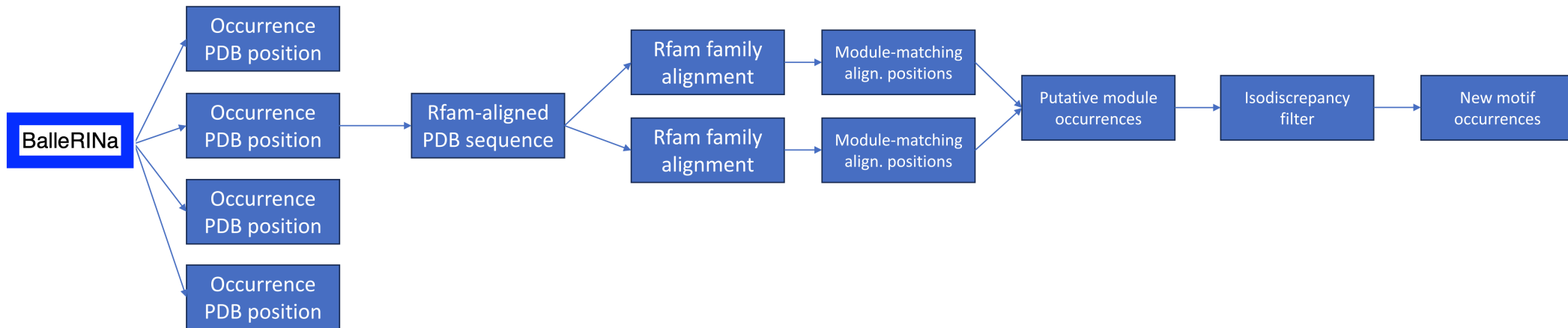
3-way junction of the c-di-amp family (RF00379)

Evolutionary information makes the problem easier



Going back to balleRINa

Rfam allows to add sequence diversity

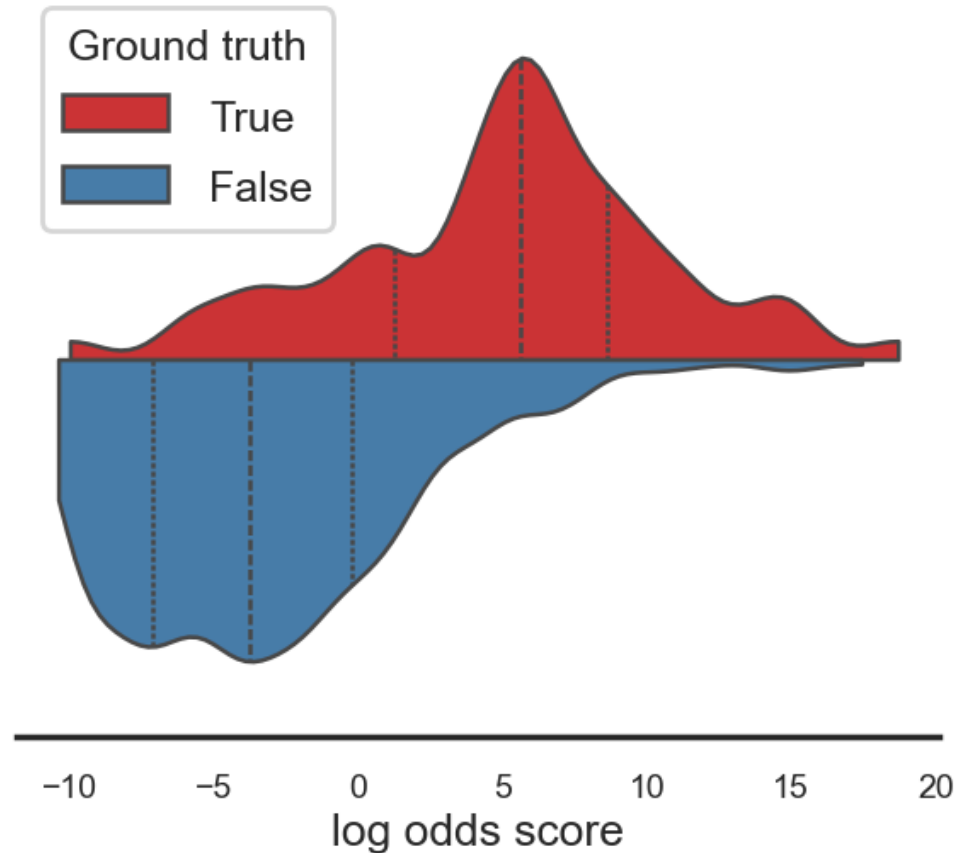


Does it work? (This is from yesterday)

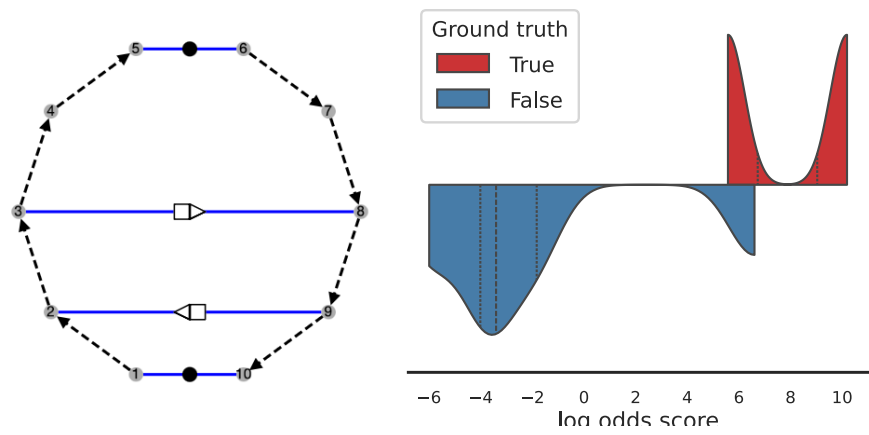
- 131 balleRINa mapped to Rfam family
- Benchmarked finished on first 101



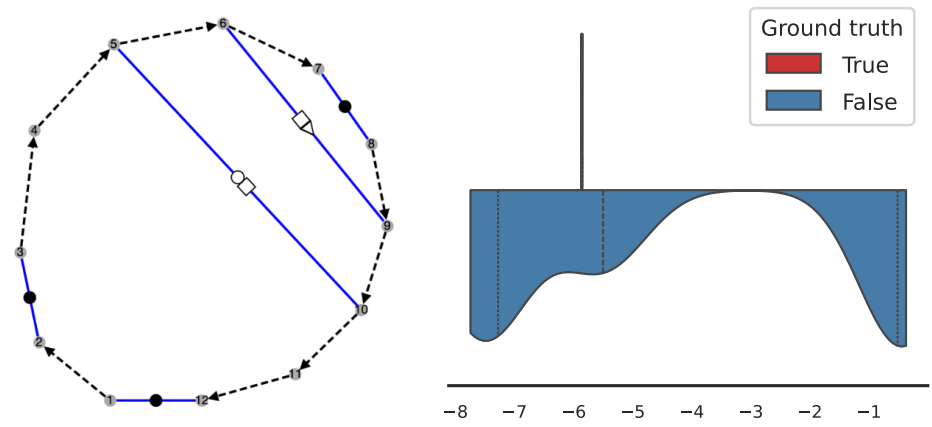
Wilfried Agbeto



Ballerina 30



Ballerina 54





Sharing is caring

<https://carnaval.cbe.uqam.ca>



[RIN](#) [Local RIN](#) [Ballerina](#) [Draw](#) [RNA](#) [Residues](#) [2D Models](#) [Filters](#) [API](#) [About](#)

CaRNAval: Catalog of Recurrent Interaction Networks (RINs) in all RNA molecules



RIN

Catalog of RINs annotated in all RNA structures

[► View RIN catalog](#)



Local RIN

Catalog of RINs in loops annotated in all RNA structures

[► View local RIN catalog](#)



Ballerina

Bulge-Agnostic Local Loop Extending RIN Annotation

[► View Ballerina catalog](#)



Draw

Draw or describe an interaction network to find in RINs or RNAs.

[► Draw an interaction network](#)



RNA

For any RNA structure all RIN annotations in 2D and 3D.

[► View RNA annotations](#)

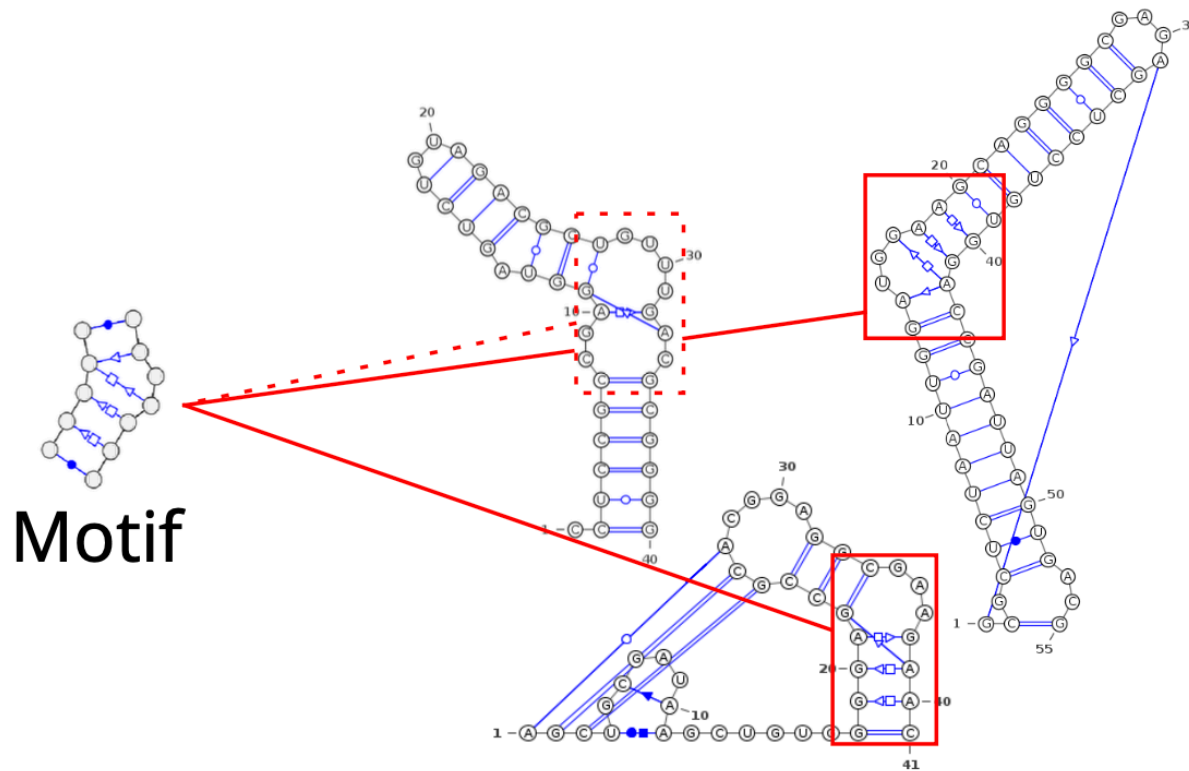


Residues

All residue types with the RINs they are participating in.

[► View all residues](#)

We can also do fuzzy graph matching
Another challenge to organize



Subparts of RNA 4V9F

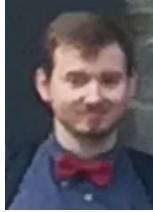
And next

- How can we extend the framework to predict modules spanning multiple loops?
- How can we group RINs / BalleRINas into functional families?
- Do some RINs correlate with inter-RNA interactions, or proteins?
- How can we identify (automatically) all Ribosome (or other type) of chains in the PDB?
- How can we use that knowledge for full 3D prediction? For design?

Thanks to all



Wilfried Agbeto



Théo Boury

Roman Sarrazin-Gendron



Hua-Ting Yao



Camille Coti

Jérôme Waldispühl



Yann Ponty



Alvar Nuñez Cabeza de Vaca



*Fonds de recherche
Santé*

Québec



INNOVATION

Fondation canadienne
pour l'innovation

Canada Foundation
for Innovation



Digital Research
Alliance of Canada

Alliance de recherche
numérique du Canada



Recurrent Interaction networks



- find all the Recurrent Interaction Networks (global and local)
- bring all of them together in carnaval.cbe.uqam.ca
- bind them all to Bulge Agnostic Local Loop Extended RIN Annotation (balleRINa)

BayesPairing2: Finding all subgraphs isomorphisms

<https://jwgitlab.cs.mcgill.ca/sarrazin/rnabayespairing2>

Pasigraph: Finding all subgraphs isomorphisms

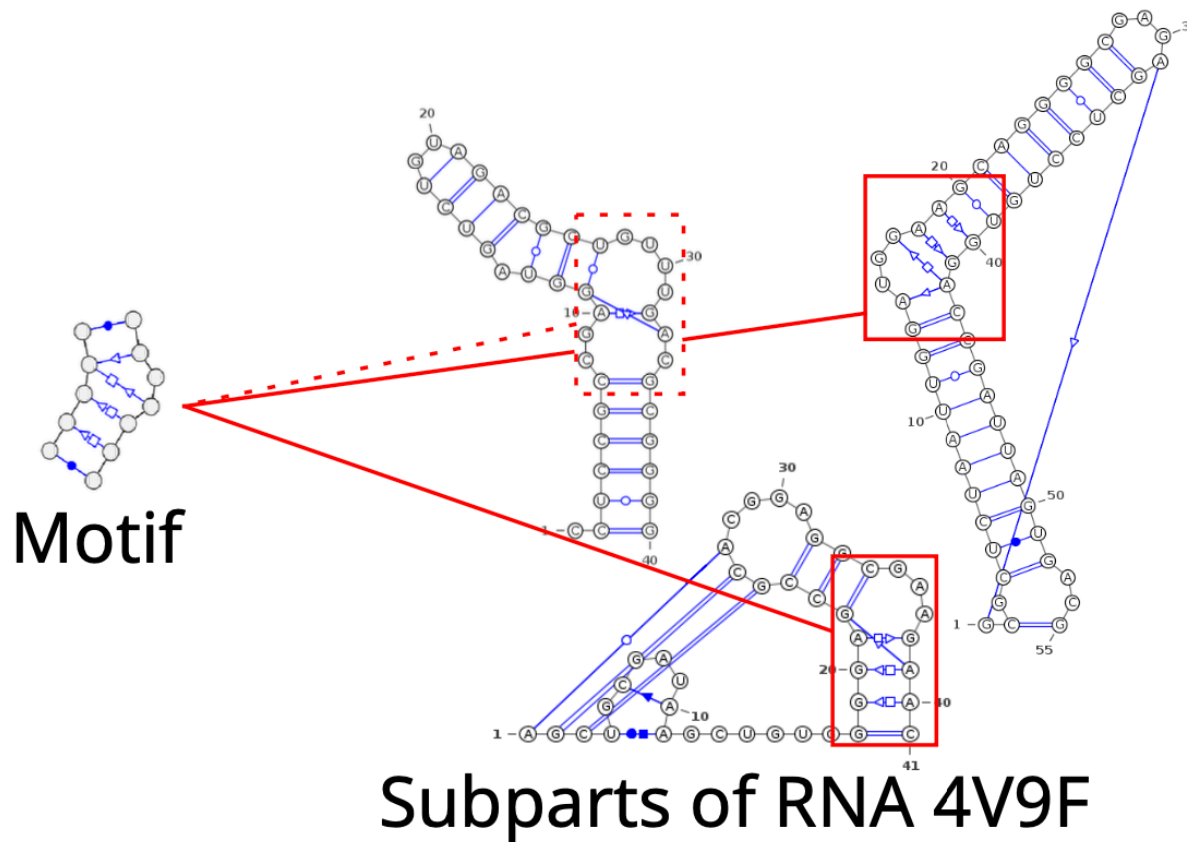
<https://gitlab.info.uqam.ca/cbe/pasigraph>

FuzzTree: Exploring module variants with explicit control of distance

<https://github.com/theoboury/FuzzTree>

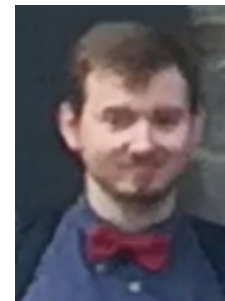
BalleRINa: Bulge Agnostic Local Loop Extended RIN Annotation

Can we do fuzzy graph matching?

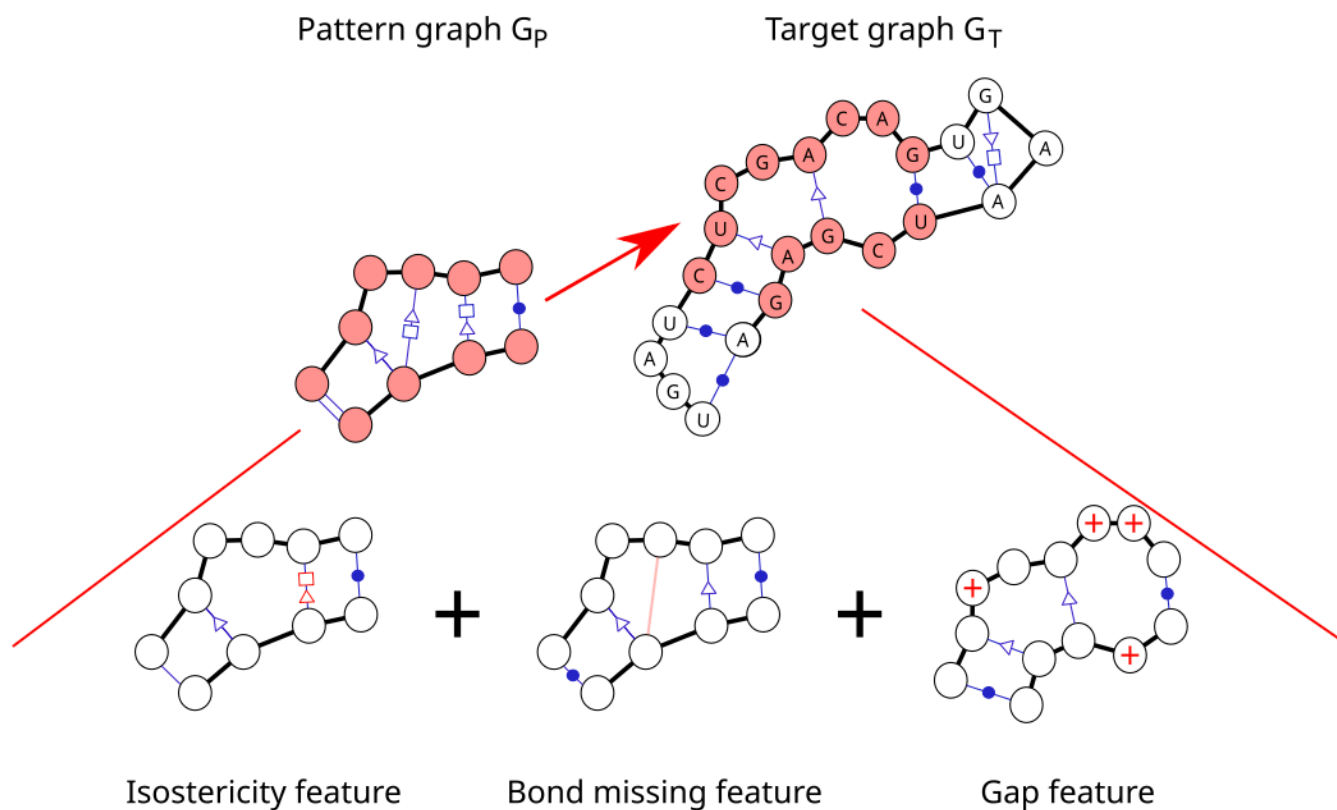


FuzzTree

github.com/theoboury/FuzzTree (wabi 2024)



Théo Boury



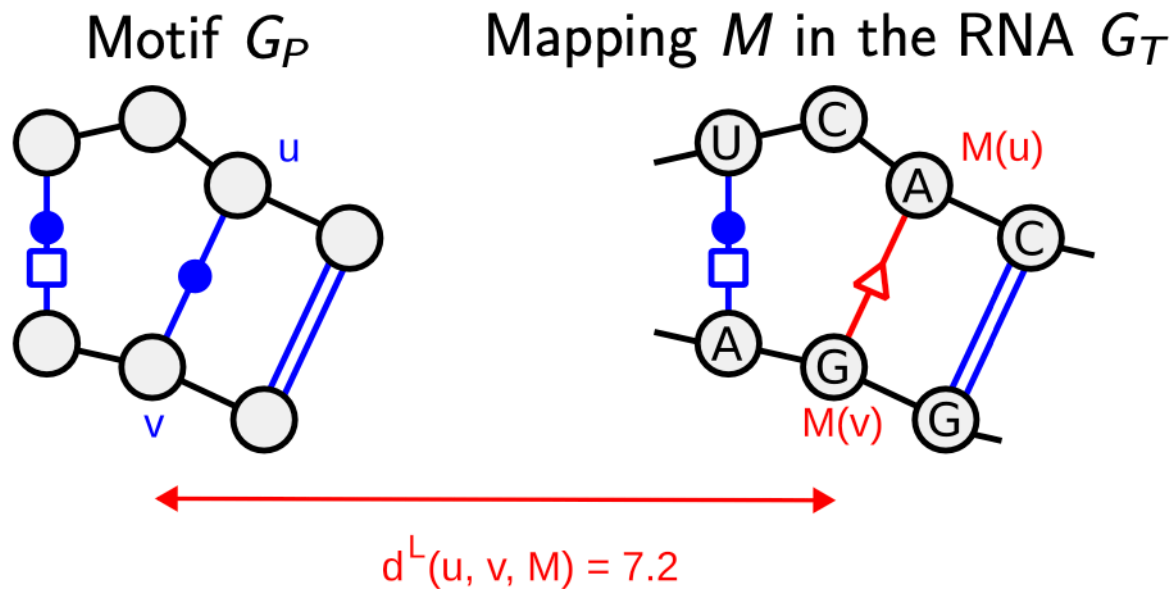
Boltzmann sampling of graphs

For each graph mapping define an energy:

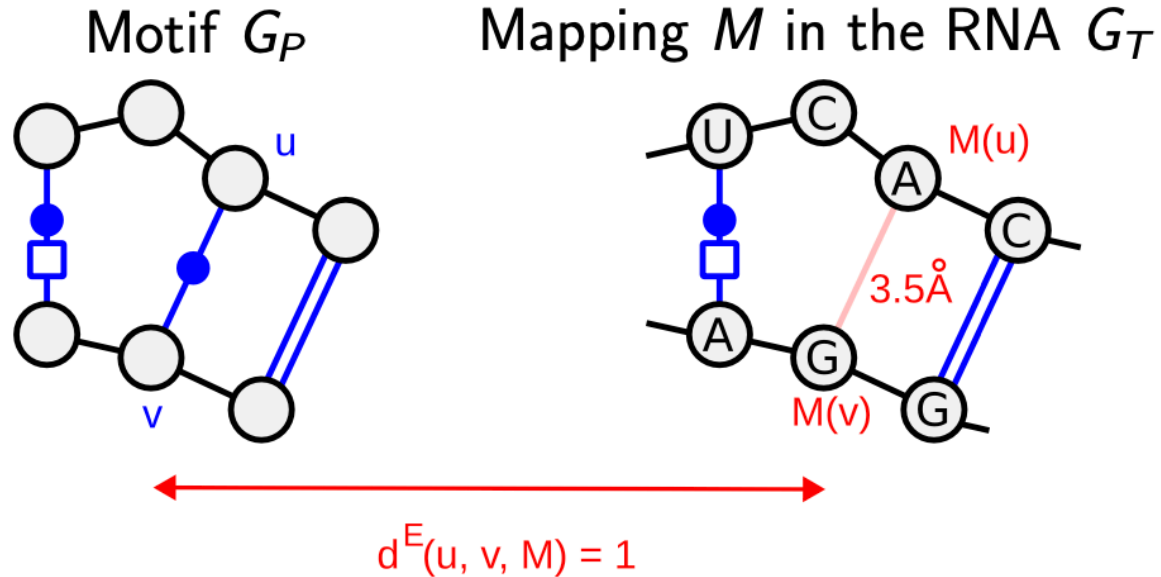
$$E(M) = \sum_{(u,v) \in E_P} \underbrace{w_L \times d^L(u, v, M)}_{\text{BP isostericity}} + \underbrace{w_E \times d^E(u, v, M)}_{\text{Missing edges}} + \underbrace{w_G \times d^G(u, v, M)}_{\text{Inserted nodes}}$$

Probability of sampling a mapping: $\mathbb{P}(M) = \frac{e^{-E(M)}}{\mathcal{Z}}$ where $\mathcal{Z} = \sum_{M'} e^{-E(M')}$

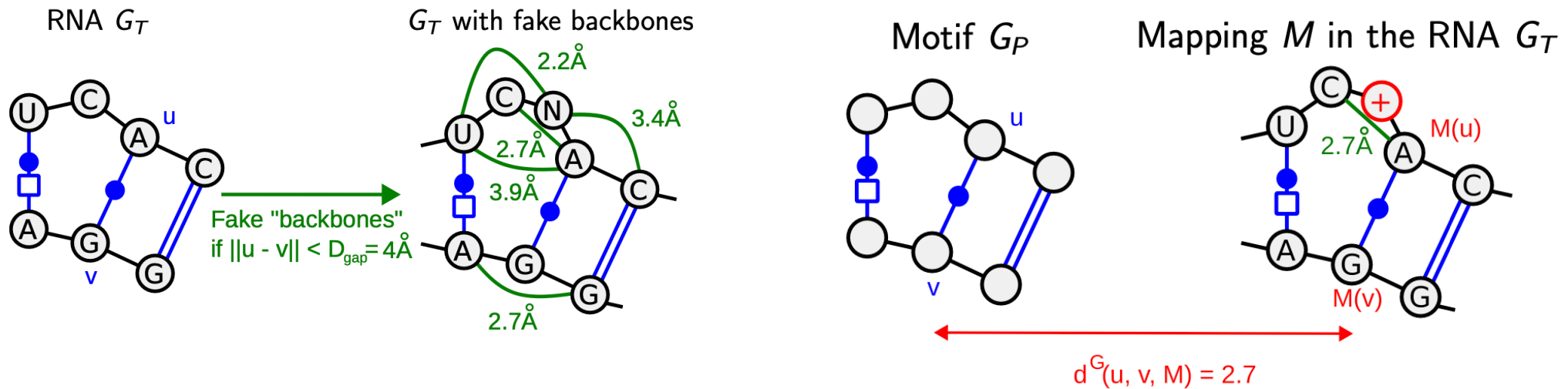
Isostericity distance (1/3)



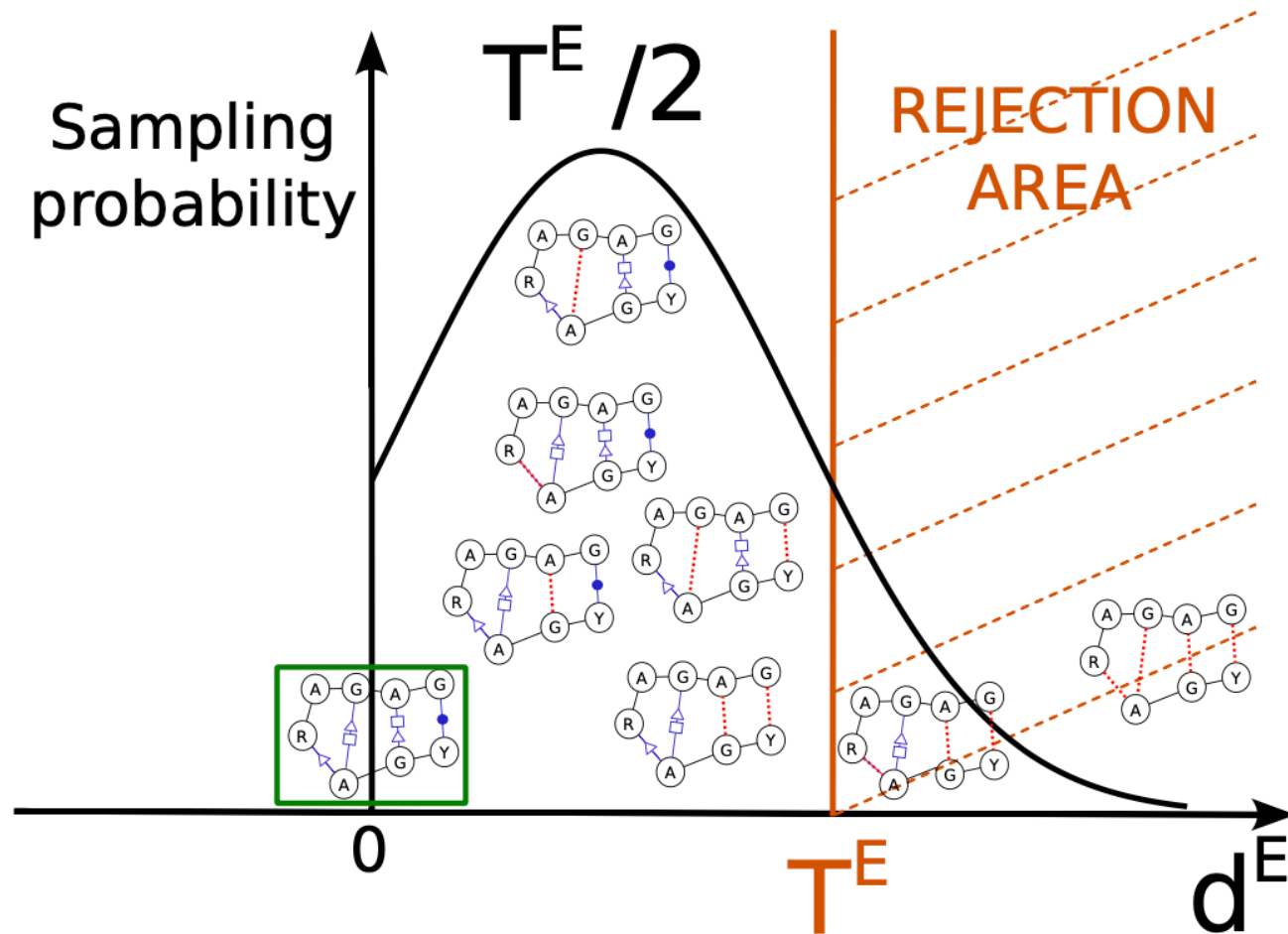
Missing edges (2/3)



Inserted nodes (3/3)



Parametrizing the sampling



Complexity exponential in tree width of pattern

Computing partition function: $O(kn^{\phi+1})$

Sampling: $O(knt)$

n : nodes target

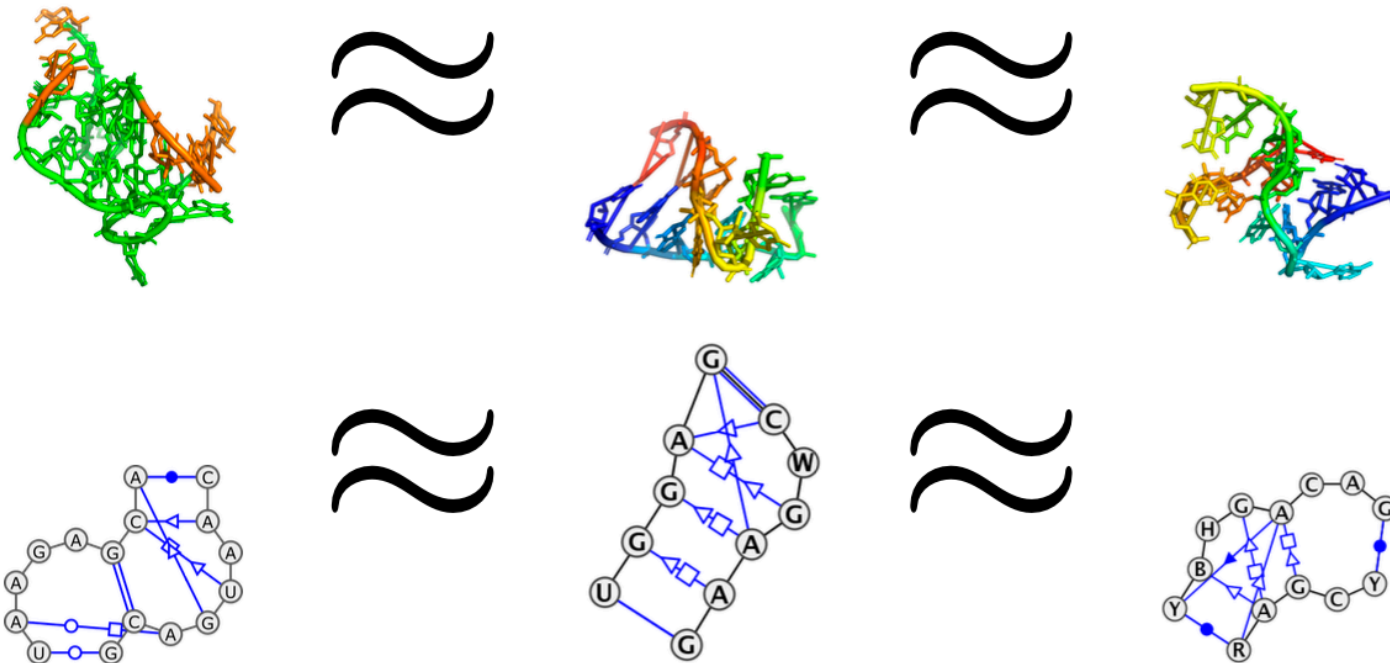
k : nodes pattern

ϕ : treewidth pattern

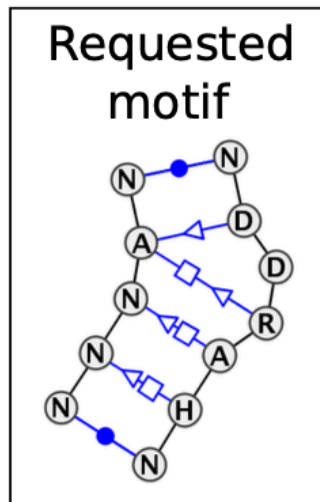
t : number samples

Back to kink-turn from RNA3DHub

- A biological family that contains 72 known motifs over more than 25 different RNAs.
- Kink-Turns are clustered in 18 different families according to atomic cristallography. ⁴

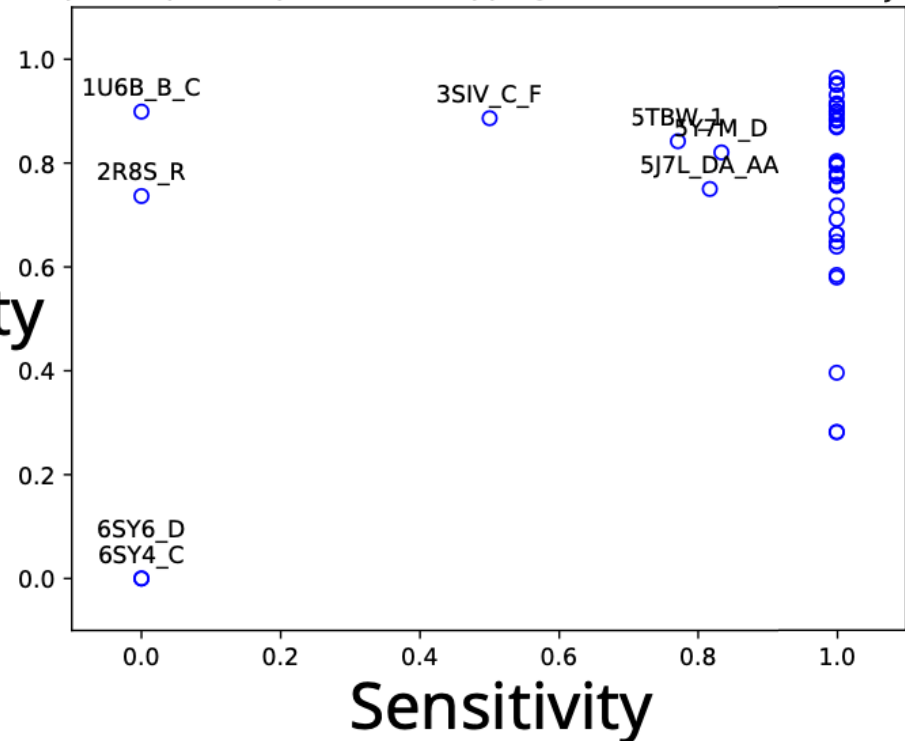


From one pattern we find most occurrences

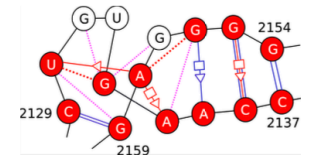
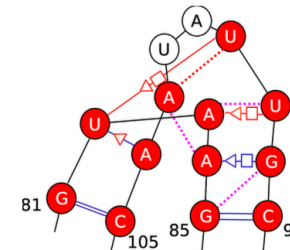
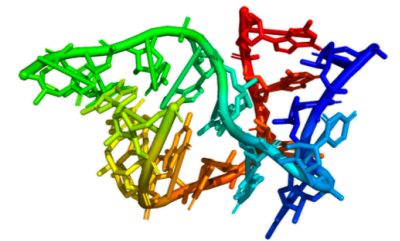
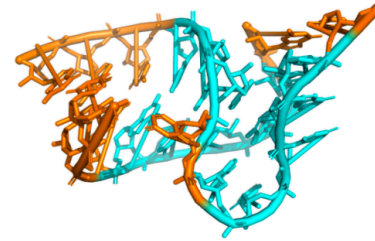
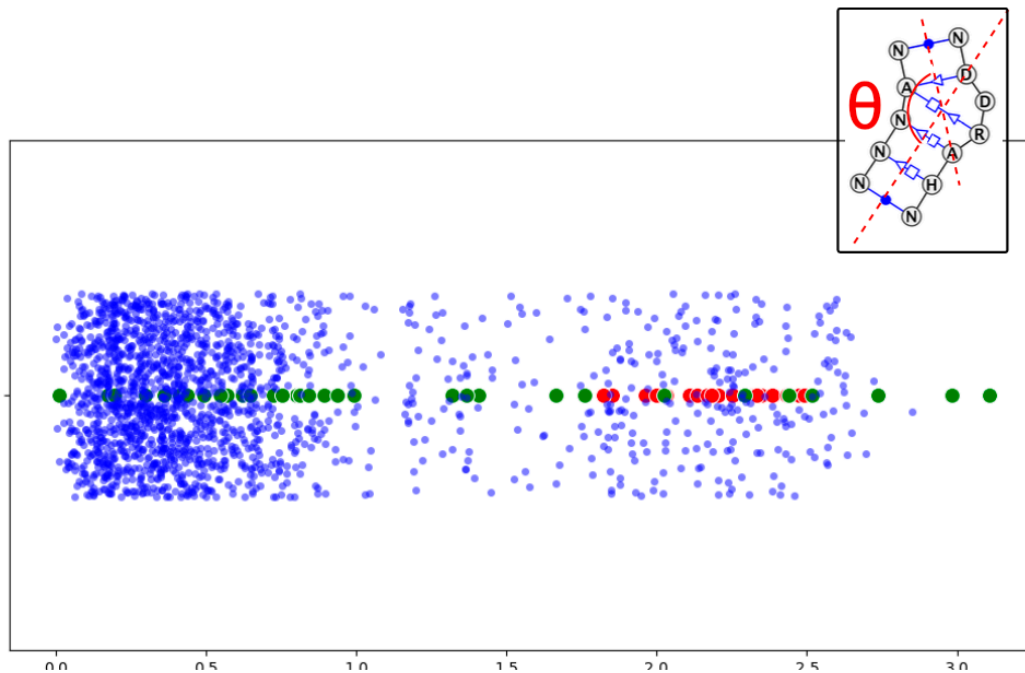


Specificity

Sensitivity and specificity of found mappings for the Kink Turn family with near

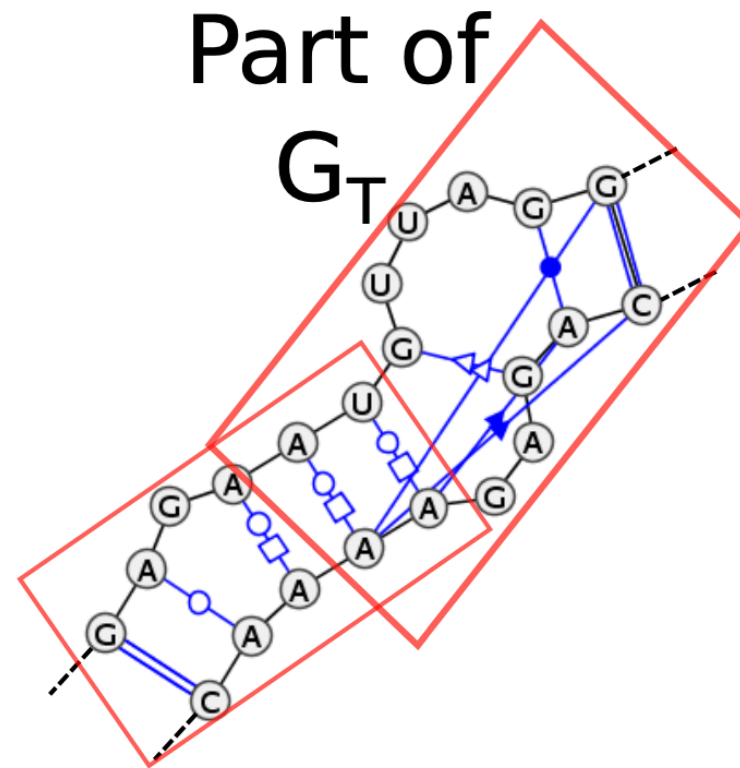
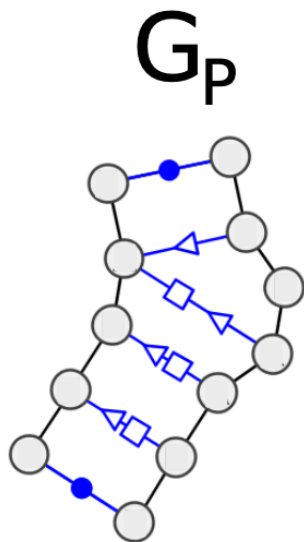


New candidates appear

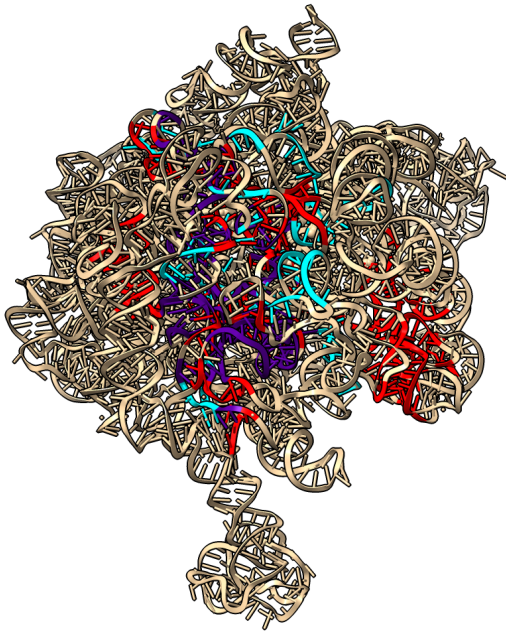


Designing core patterns a challenge

A kink-turn that can not be found

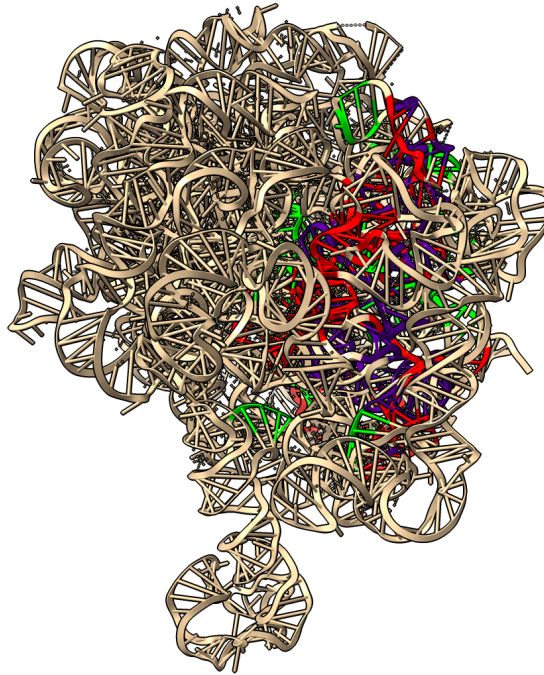


Large interaction networks span many loops



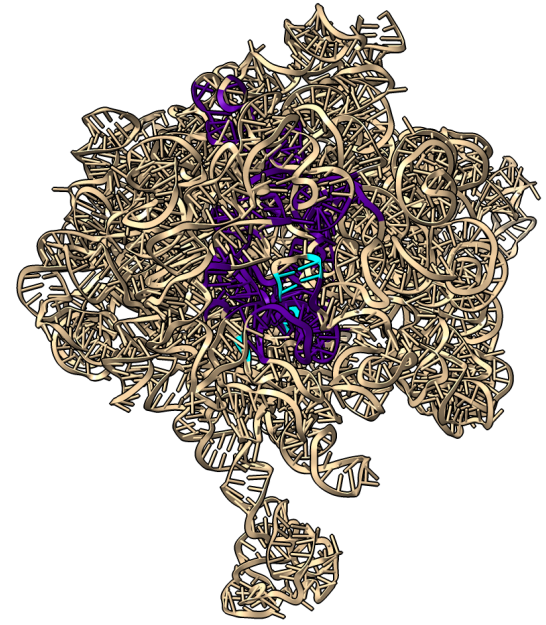
5J7L

E. Coli



4Y4O

T. Thermophilus



6SPB

P. aeruginosa

Experiment	n sequences
Putative C/D box snoRNA sequences identified in snoDB	1223
CacoFold hits for KT on sequences searched	740
CacoFold hits overlapping with predicted C/D or C'/D'	135
BP2.5 hits for KT on sequences searched	842
BP2.5 hits overlapping with predicted C/D or C'/D'	733
BP2.5 overlapping hits with score > 0 (low confidence)	693
BP2.5 overlapping hits with score > 2 (med. confidence)	524
BP2.5 overlapping hits with score > 5 (high confidence)	282