Finding entanglements of RNA structural elements

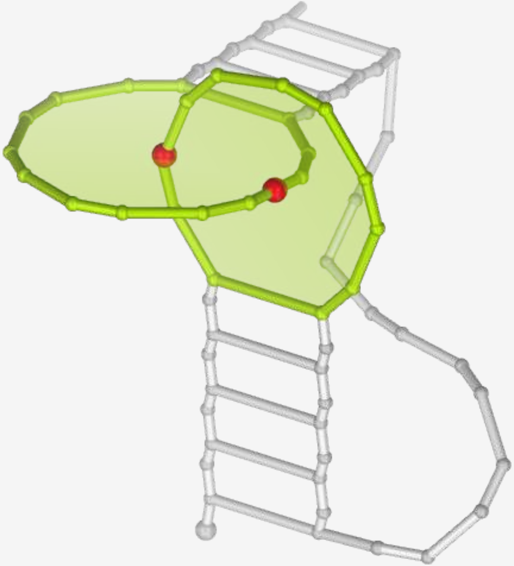
2022-08-09

Tomasz Zok

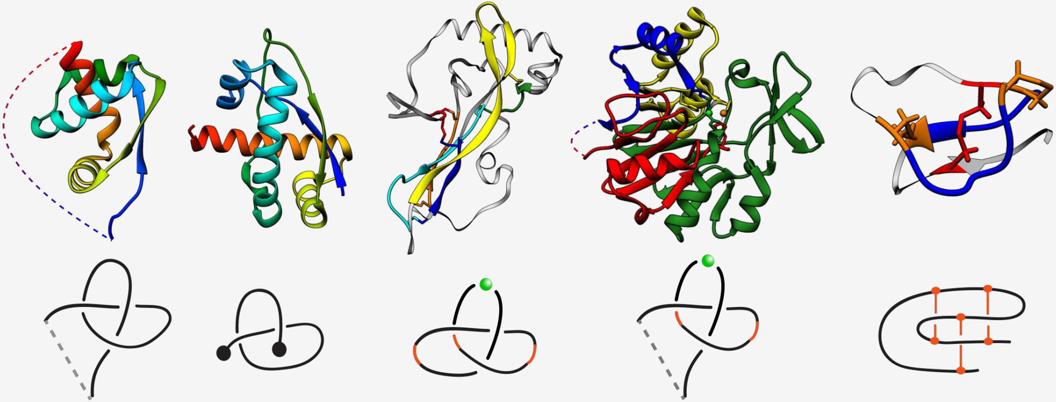
Institute of Computing Science, Poznan University of Technology

Introduction

- Entanglements of structural elements are motifs / phenomena found in RNA structures
- In some cases they are actually desired (i.e., functional), in others they are likely a modeller mistake or a computational artifact
- We:
  - defined these motifs and their classification,
  - designed algorithms to find them,
  - evaluated the findings on different datasets, and
  - prepared a user-friendly application



Disclaimer: protein knots



Probabilistic knot

Knotoid

Deterministic knot including disulfide

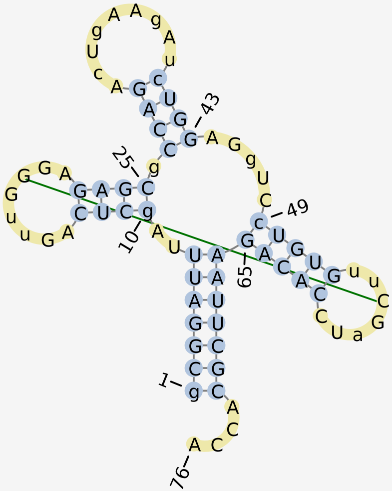
Probabilistic knot including disulfide

Cysteine "knot"

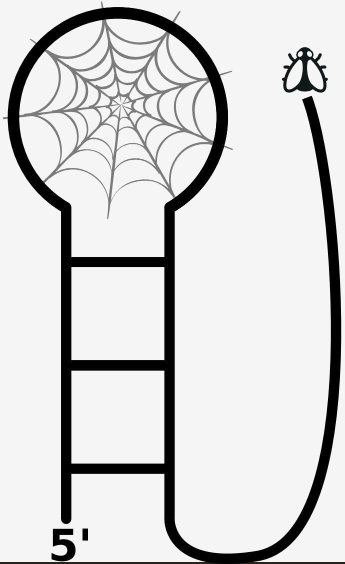
KnotProt 2.0: A Database of Proteins with Knots and Other Entangled Structures. P. Dabrowski-Tumanski et al. Nucleic Acids Research. 2019. 47(D1):D367–D375

RNA structure

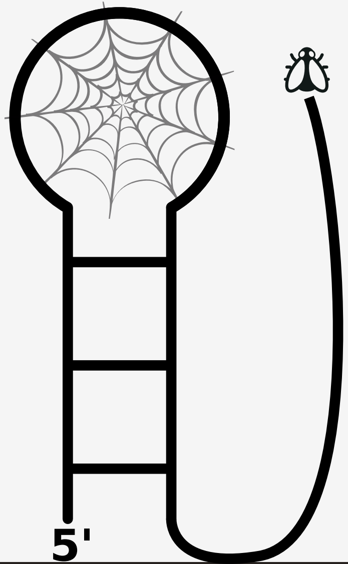
- We look at RNA structures as a composition of structural elements:
    - Double strands / Duplexes / Stems
    - Hairpin loops
    - Bulges
    - Internal loops
    - Multiloops / N-way junctions
    - Single strands
    - Pseudoknots

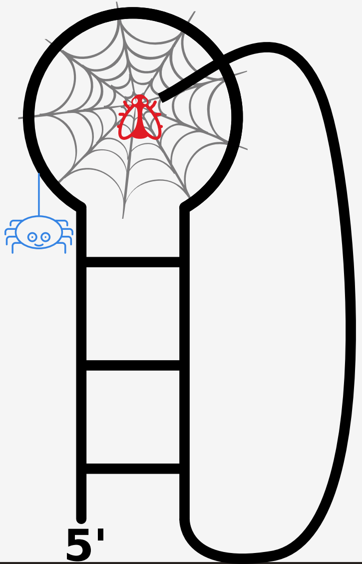


RNAspider idea



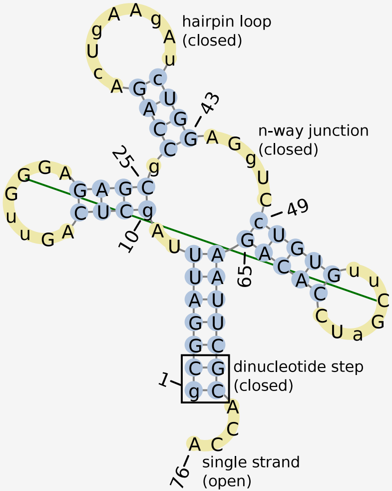
RNAspider idea





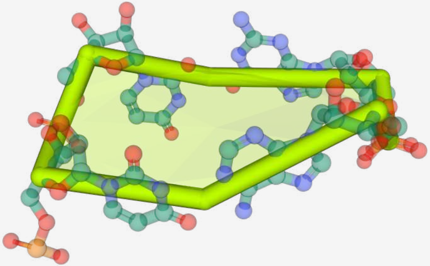
Open and closed elements

- In our approach, we look at the elements in a slightly different way:
    - Single strands are called open elements
    - Loops are closed elements
    - Each two consecutive base pairs in a stem (named a *dinucleotide step*) is also a separate closed element



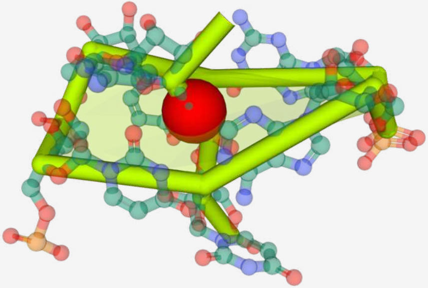
Polygons and meshes

- RNAspider constructs a polygonal chain by connecting selected atoms and pseudoatoms (geometrical centers of base pairs)
- For closed elements, this polygonal chain will reach back to the first point
- This allows to fill the closed area with polygon mesh via triangulation

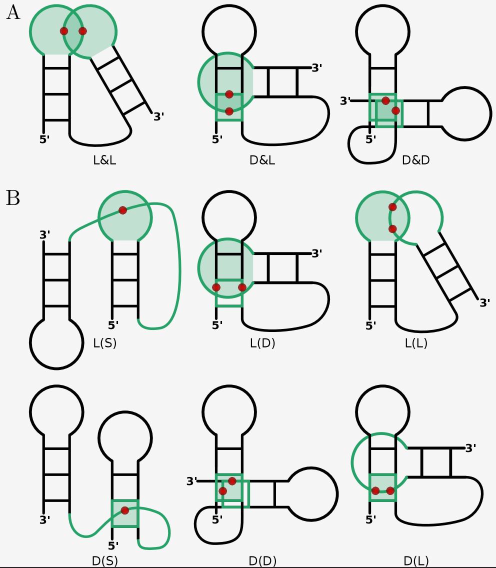


Polygons and meshes

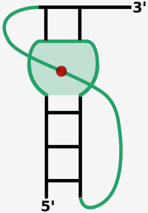
- We employ Möller-Trumbore algorithm to detect intersections between triangles and lines sections (fragments of the polygonal chain)
- This allows to distinguish between punctured and puncturing structural element
- This also provides exact 3D coordinates of the point of intersection



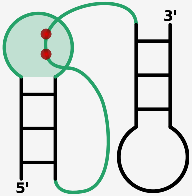
Entanglements classification



Entanglements classification

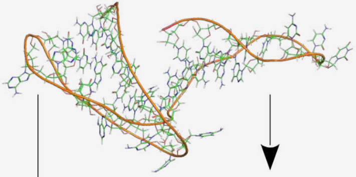


higher-order L(S)



L(S.)

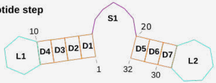
1. Input RNA 3D structure model



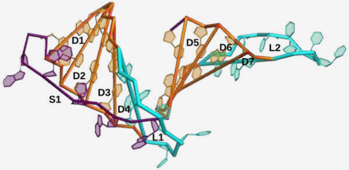
2. Extract the secondary structure

```
GGCAUCUCCGUGCCAAUAAGGCCCUCUCGGCC
(((((....))))).....((((.....))))
```

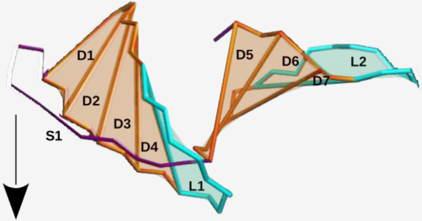
D – dinucleotide step
L – loop
S – strand



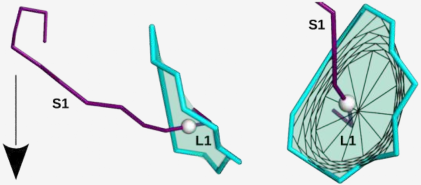
3. Divide input 3D structure into structure elements based on the 2D topology



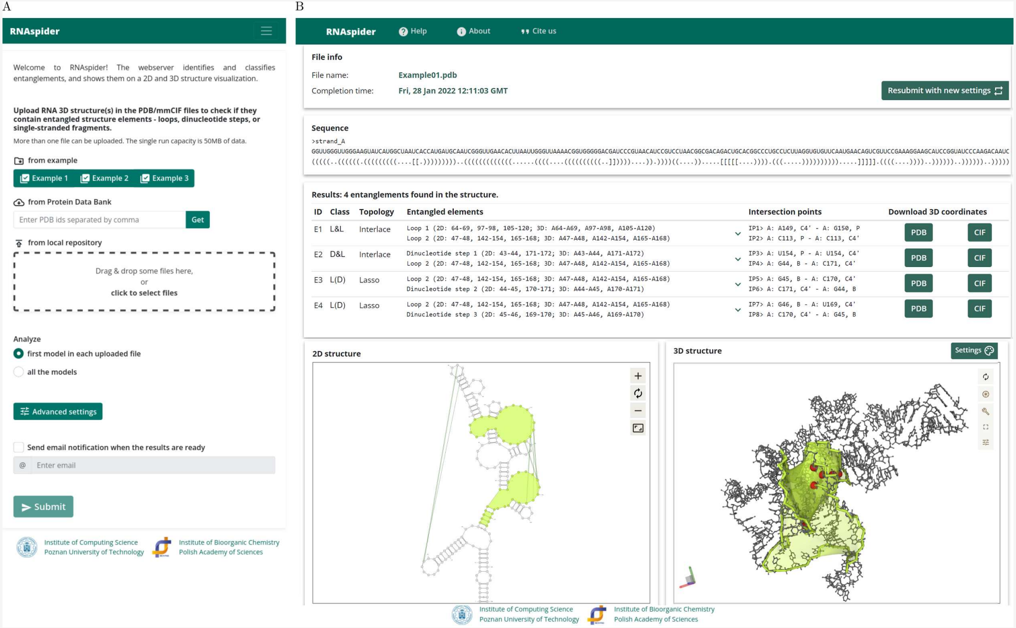
4. For each encircled 3D structure element, perform the triangulation of its inner area



5. Apply the segment intersection algorithm to every triangle of the polygon mesh. Detect punctures and identify entanglements.



6. Output the list of entanglements with the classification



RNAspider

Welcome to RNAspider! The webserver identifies and classifies entanglements, and shows them on a 2D and 3D structure visualization.

Upload RNA 3D structure(s) in the PDB/mmCIF files to check if they contain entangled structure elements - loops, dinucleotide steps, or single-stranded fragments.

More than one file can be uploaded. The single run capacity is 50MB of data.

from example

☑ Example 1 ☑ Example 2 ☑ Example 3

from Protein Data Bank

Enter PDB ids separated by comma [Get]

from local repository

Drag & drop some files here,
or
click to select files

Analyze

● first model in each uploaded file

○ all the models

[Advanced settings]

☐ Send email notification when the results are ready

@ Enter email

[Submit]

Institute of Computing Science
Poznan University of Technology

Institute of Bioorganic Chemistry
Polish Academy of Sciences

RNAspider Help About Cite us

File info

File name: **Example01.pdb**

Completion time: **Fri, 28 Jan 2022 12:11:03 GMT**

[Resubmit with new settings]

Sequence

```
>strand_A
GGUUGGGUUGGGAAGUAUCAUGGCUAAUCACCAUGAUGCAAUCGGGUUGAACACUUAAUUGGGUUAAAACGGUGGGGGACGAUCCCGUAACAUCCGUCCUAACGGCGACAGACUGCACGGCCCUGCCUCUUAGGUGUGUUCAAUGAACAGUCGUUCCGAAAGGAAGCAUCCGGUAUCCCAAGACAAUC
((((..(((((.((((((((....[[.)))))))))..((((((((((((......((((....((((((((((..]])))))....)).))))((....)).....[[[[[....)))).(((.....))))))))))).....]]]]].((((....))))..))))))..))))))..)))))
```

Results: 4 entanglements found in the structure.

ID	Class	Topology	Entangled elements	Intersection points	Download 3D coordinates	
E1	L&L	Interlace	Loop 1 (2D: 64-69, 97-98, 105-120; 3D: A64-A69, A97-A98, A105-A120) Loop 2 (2D: 47-48, 142-154, 165-168; 3D: A47-A48, A142-A154, A165-A168)	IP1> A: A149, C4' - A: G150, P IP2> A: C113, P - A: C113, C4'	PDB	CIF
E2	D&L	Interlace	Dinucleotide step 1 (2D: 43-44, 171-172; 3D: A43-A44, A171-A172) Loop 2 (2D: 47-48, 142-154, 165-168; 3D: A47-A48, A142-A154, A165-A168)	IP3> A: U154, P - A: U154, C4' IP4> A: G44, B - A: C171, C4'	PDB	CIF
E3	L(D)	Lasso	Loop 2 (2D: 47-48, 142-154, 165-168; 3D: A47-A48, A142-A154, A165-A168) Dinucleotide step 2 (2D: 44-45, 170-171; 3D: A44-A45, A170-A171)	IP5> A: G45, B - A: C170, C4' IP6> A: C171, C4' - A: G44, B	PDB	CIF
E4	L(D)	Lasso	Loop 2 (2D: 47-48, 142-154, 165-168; 3D: A47-A48, A142-A154, A165-A168) Dinucleotide step 3 (2D: 45-46, 169-170; 3D: A45-A46, A169-A170)	IP7> A: G46, B - A: U169, C4' IP8> A: C170, C4' - A: G45, B	PDB	CIF

2D structure

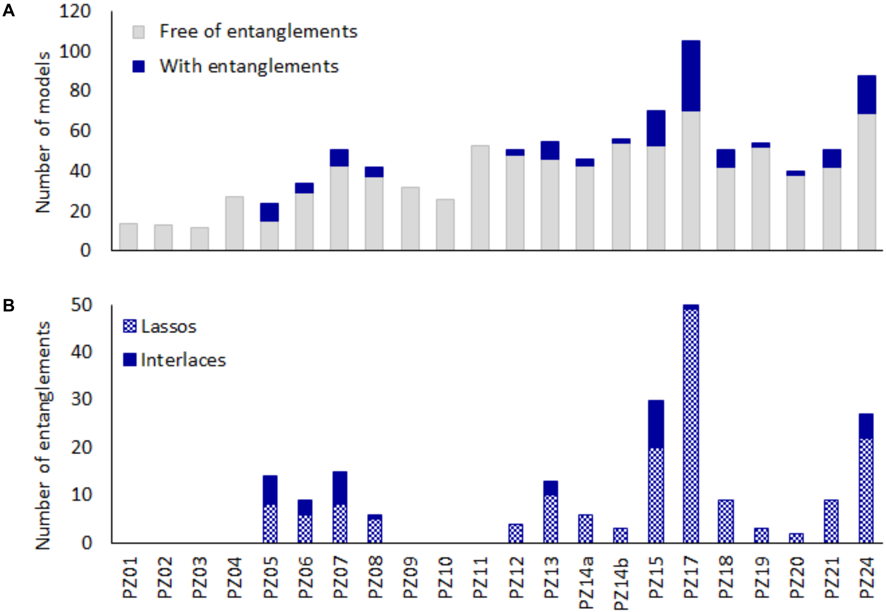
3D structure

Institute of Computing Science
Poznan University of Technology

Institute of Bioorganic Chemistry
Polish Academy of Sciences

Entanglements in the predicted models

- We decided to check all available RNA-Puzzles 3D structures for the presence of entangled structural elements
- We tested 1 017 unique RNA 3D structures (both models and targets)
- We found entanglements in 138 instances (137 models and 1 target: PZ18, PDB 5TPY, xrRNA from Zika virus)
- Most entangled models contained just one entanglement (usually of lasso topology)
- All together, we found 201 entanglements - 165 lassos and 36 interlaces

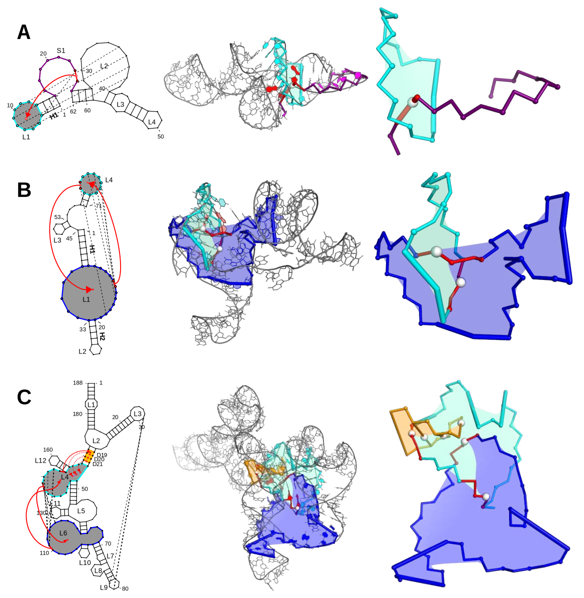

A
120
100
80
60
40
20
0
Number of models
Free of entanglements
With entanglements
B
50
40
30
20
10
0
Number of entanglements
Lassos
Interlaces
PZ01
PZ02
PZ03
PZ04
PZ05
PZ06
PZ07
PZ08
PZ09
PZ10
PZ11
PZ12
PZ13
PZ14a
PZ14b
PZ15
PZ17
PZ18
PZ19
PZ20
PZ21
PZ24

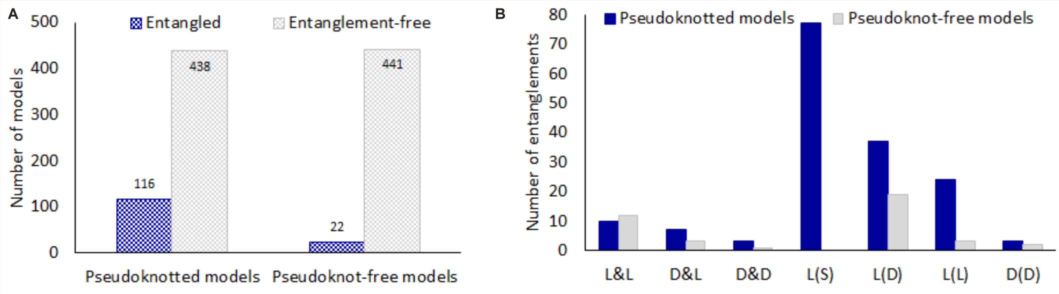
Puzzle	Length	RNA type	Ent. models	Entanglements
PZ17	62	Pistol ribozyme	35 / 105	1xL&L 32xL(S) 9xL(D) 8xL(L)
PZ24	112	Viral non-coding RNA	19 / 88	2xL&L 3xD&L 10xL(S) 9xL(D) 3xL(L)
PZ15	68	Hammerhead ribozyme	17 / 70	10xL&L 5xL(S) 14xL(D) 1xL(L)

A) 7th position according to RMSD in PZ17 (Pistol ribozyme), L(S) entanglement

B) 43rd position according to RMSD in PZ13 (ZTP riboswitch), L&L entanglement

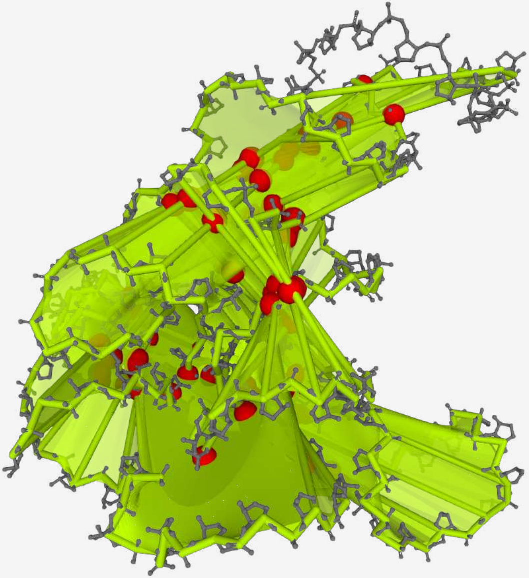
C) 3rd position according to RMSD in PZ5 (lariat-capping ribozyme), L&L, L&D, 2xL(D)

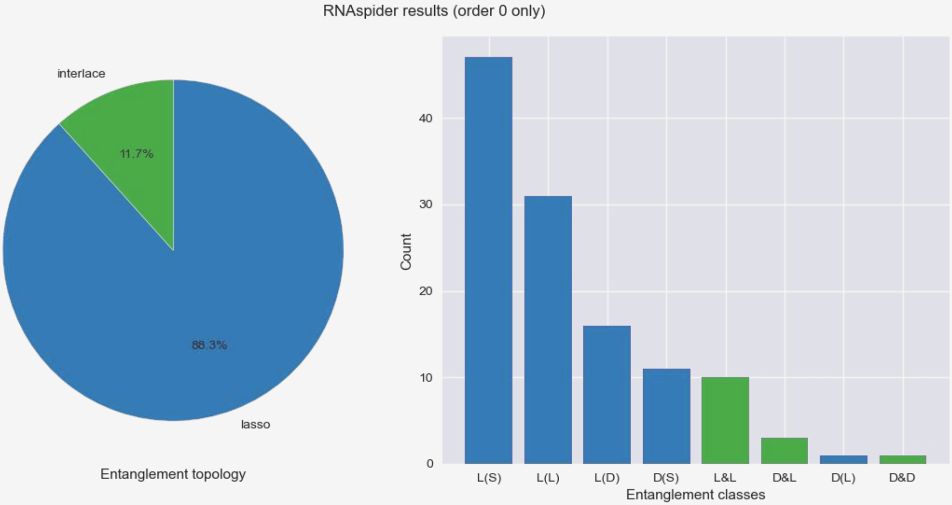



A
500
400
300
200
100
0
Number of models
▨ Entangled
▨ Entanglement-free
116
438
22
441
Pseudoknotted models
Pseudoknot-free models
B
80
70
60
50
40
30
20
10
0
Number of entanglements
■ Pseudoknotted models
■ Pseudoknot-free models
L&L
D&L
D&D
L(S)
L(D)
L(L)
D(D)

Entanglements in the PDB

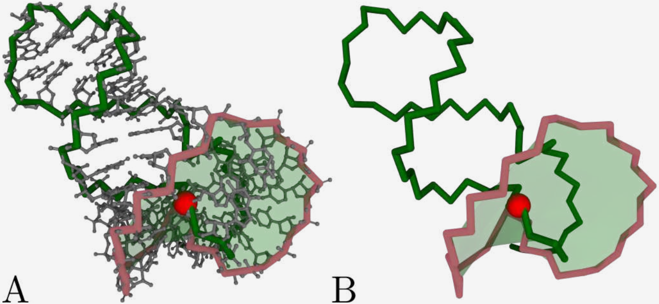
- We also decided to analyze all available 5 804 RNA 3D structures solved experimentally
- Interestingly, 11 entries contained overlapping/clashing atoms so our tool reported insane amounts of entanglements
- From the remaining cases, we found 120 primary entanglements and 3 308 higher-order ones




RNAspider results (order 0 only)
interlace
11.7%
88.3%
lasso
Entanglement topology
Count
40
30
20
10
0
L(S)
L(L)
L(D)
D(S)
L&L
D&L
D(L)
D&D
Entanglement classes

Knot-like fold

- Exoribonuclease-resistant RNAs (xrRNAs) are found in viruses – so far found in *flaviviruses* (dengue, yellow fever, West Nile, Zika) and in SARS-CoV-2 FSE
- Their structure has properties which protects them from being cut by the enzyme
- It turns out, xrRNAs have a higher-order entanglement (in the literature named a *knot-like fold* or a *ring-like architecture*)


A
B

Summary

- Biological molecules may form knots (proteins) or entanglements / knot-like folds (RNAs)
- The latter are defined on structural elements (loops, dinucleotide steps and single strands) where some are punctured / intersected by others
- Some entanglements of structural elements are *functional*
- Others are probably *artifacts of software processing*
- RNAspider is the first and only tool to find and classify them

Publications

1. Entanglements of Structure Elements Revealed in RNA 3D Models. M. Popenda, **T. Zok**, J. Sarzynska, A. Korpeta, R.W. Adamiak, **M. Antczak** and **M. Szachniuk**. Nucleic Acids Research. 2021. 49(17):9625–9632. doi:10.1093/nar/gkab716
2. RNAspider: A Webserver to Analyze Entanglements in RNA 3D Structures. K. Luwanski, V. Hlushchenko, M. Popenda, **T. Zok**, J. Sarzynska, D. Martsich, **M. Szachniuk** and **M. Antczak**. Nucleic Acids Research. 2022. gkac218. doi:10.1093/nar/gkac218