Misfolded or not - that is the question…
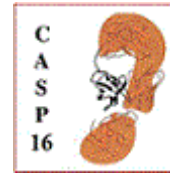
Marta Szachniuk

Poznan University of Technology, PL

Institute of Bioorganic Chemistry PAS, PL

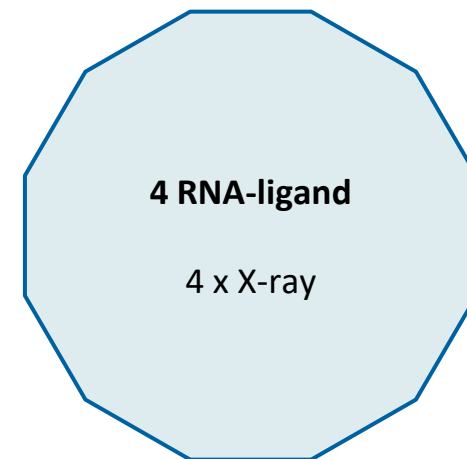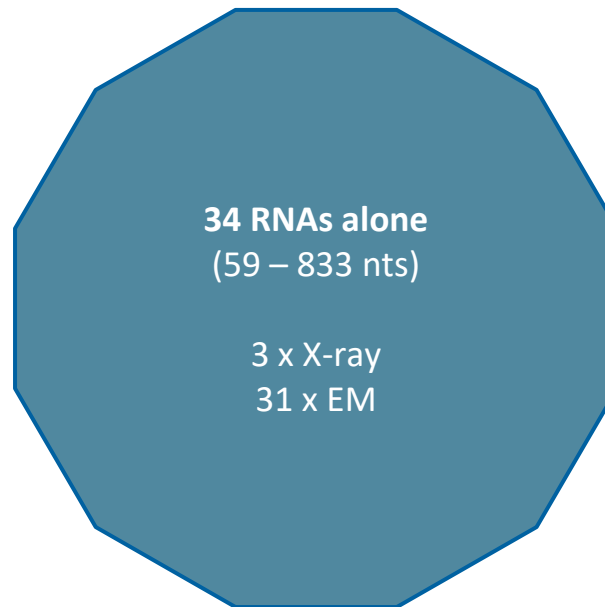Benasque, July 22, 2024

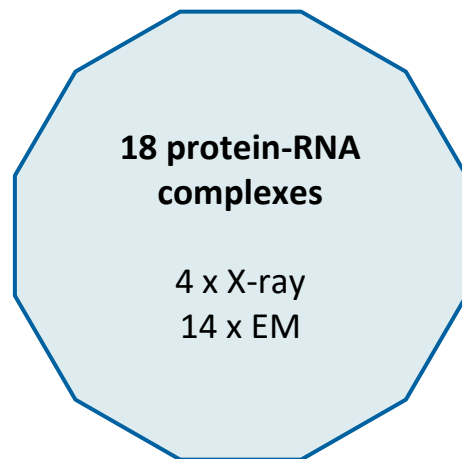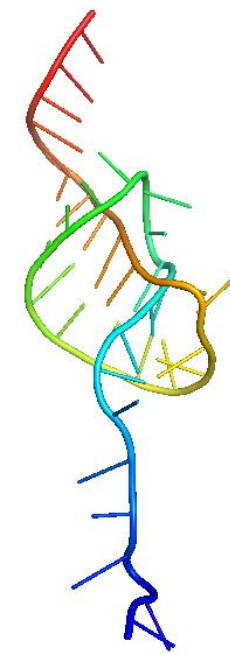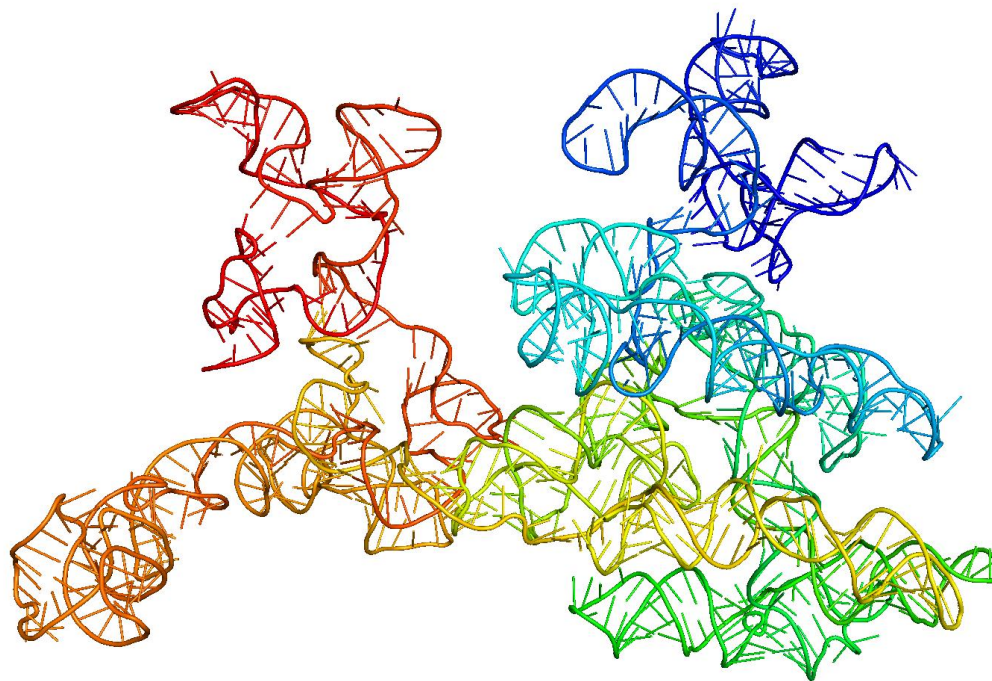# CASP16

Modeling season: 1 May – 31 August, 2024
(targets released: 1 May – 31 July, 2024)

What's behind: 12 weeks of modeling, 56 RNA targets

34 RNAs alone
(59 – 833 nts)

3 x X-ray
31 x EM

18 protein-RNA complexes

4 x X-ray
14 x EM

4 RNA-ligand
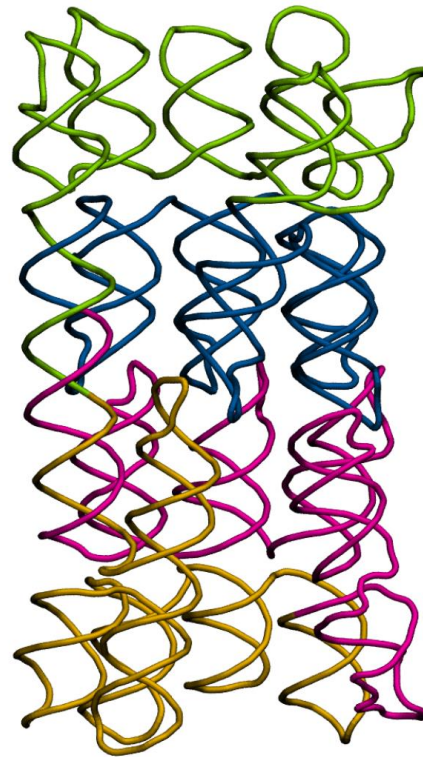
4 x X-ray

Misfolded or not?

If the native structure is unknown…
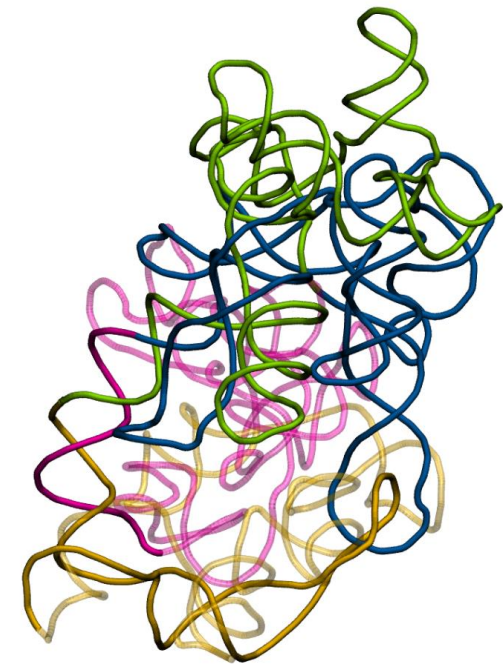
Misfolded or not?

predicted model

native structure
(720 nt)

predicted model

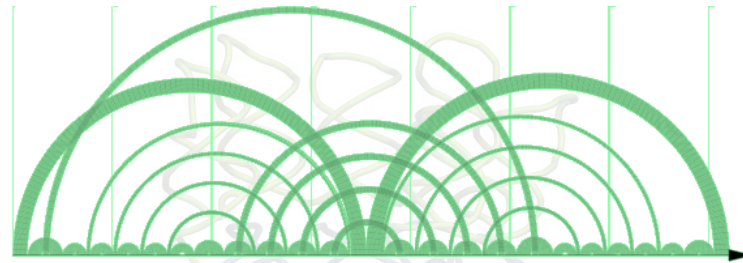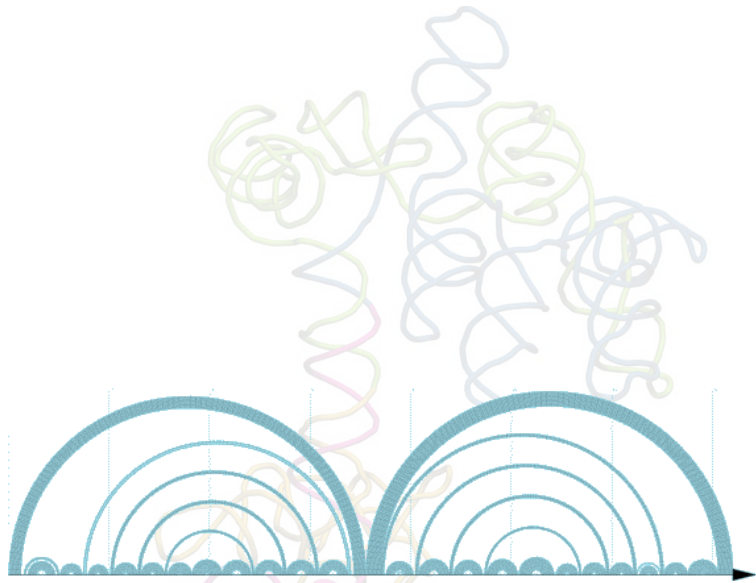# Misfolded or not?

predicted model

native structure
(720 nt)

predicted model

❌ **RMSD = 48.23Å**
❌ **TM-score = 0.204**
✔ **INF$_{all}$ = 0.82**
✔ **Clash score = 0.30**

2D structure looks (quite) fine.
What's wrong with the 3D topology?

RMSD = 39.73Å ❌
TM-score = 0.186 ❌
INF$_{all}$ = 0.85 ✔
Clash score = 4.81 ✔

Misfolded or not?

predicted model

native structure
(720 nt)
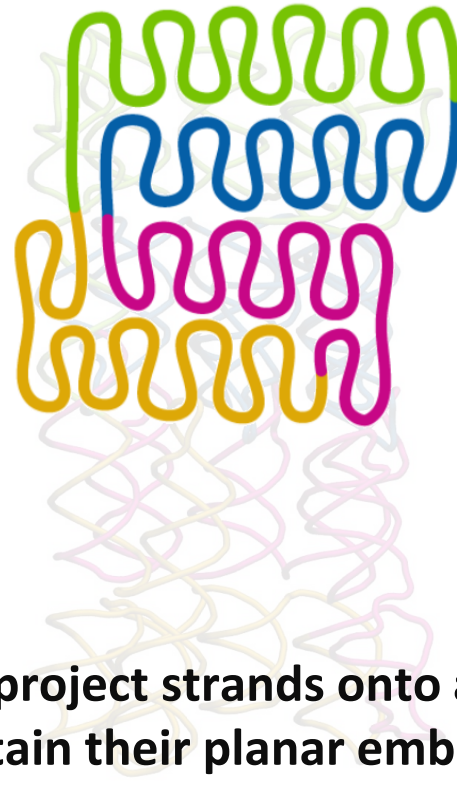
predicted model

**Let's project strands onto a plane
to obtain their planar embedding.**

✖ RMSD = 48.23Å
✖ TM-score = 0.204
✔ INF$_{all}$ = 0.82
✔ Clash score = 0.30

RMSD = 39.73Å ✖
TM-score = 0.186 ✖
INF$_{all}$ = 0.85 ✔
Clash score = 4.81 ✔

Misfolded or not?

predicted model

native structure
(720 nt)

predicted model

simplified view (to minimize twists)
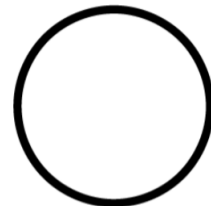
Topological knots

3_1 (trefoil)

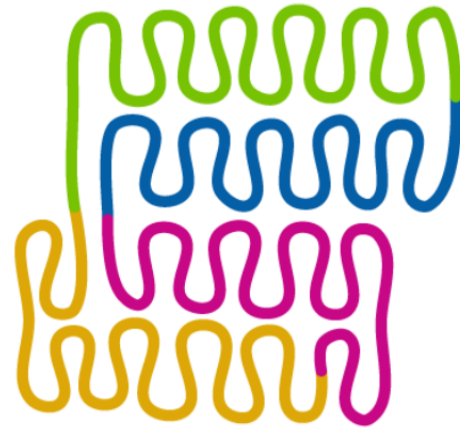0_1 (unknot)

7_2
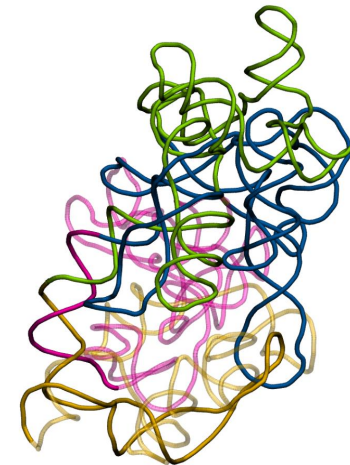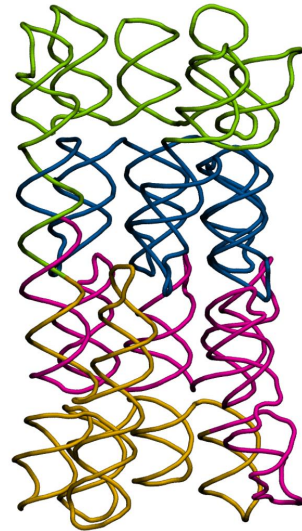
In knot theory

3_1 (trefoil)

0_1 (unknot)

genus

7_2
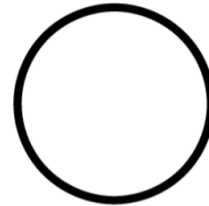
braid no.

- Topological knot – closed curve embedded in 3D space

- We used to model it with rope and to diagram as projection on a plane.

- Knot's name = genus + braid number (due to Alexander-Briggs-Rolfsen notation used here)

- Genus – crossing number

Misfolded or not?

3_1 (trefoil)

0_1 (unknot)
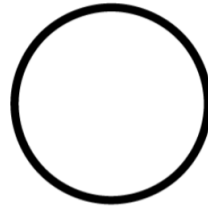
7_2

predicted model

native structure

predicted model

These predictions are misfolds!
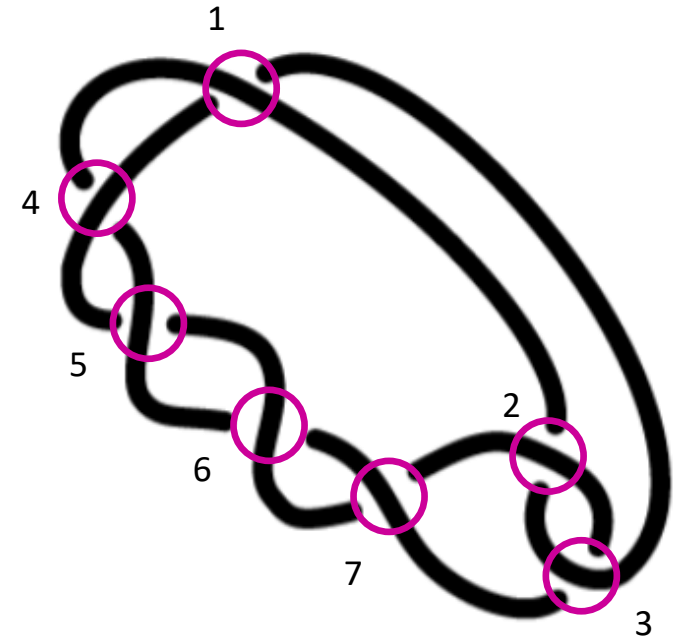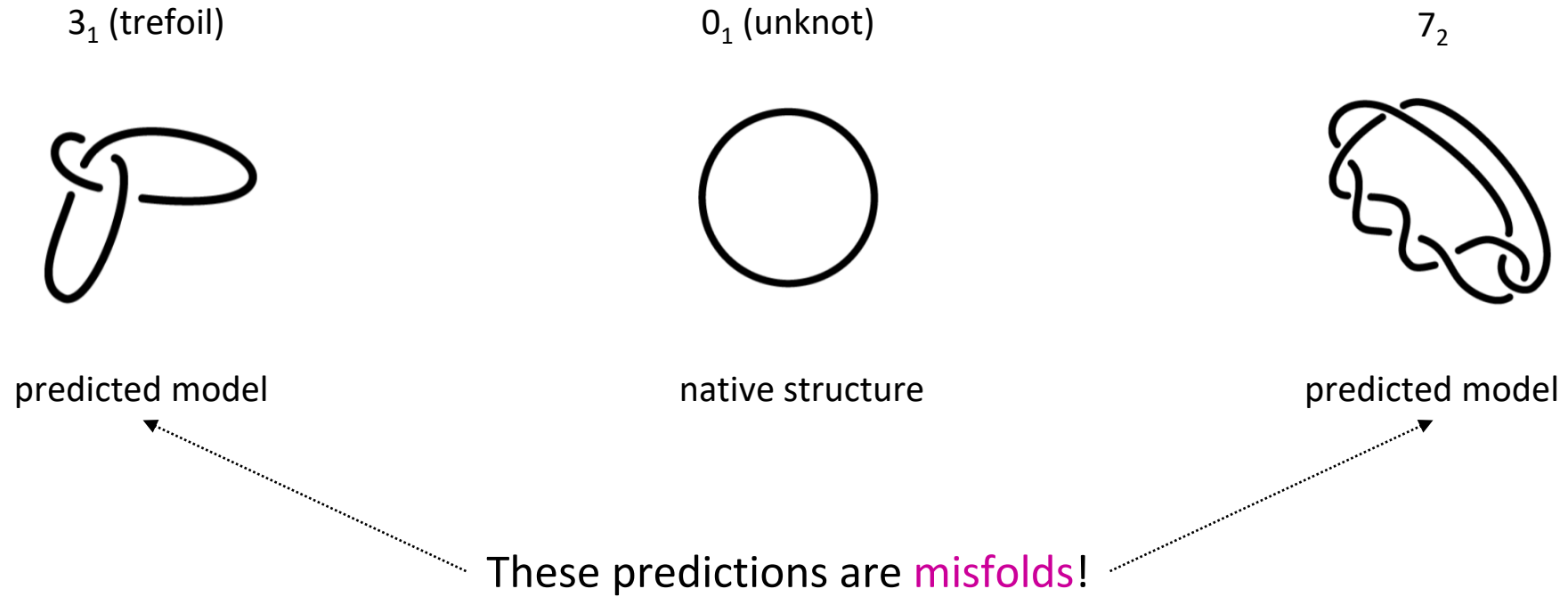
What if we don't have a reference structure?

Is a knot always a misfold?

Molecular knots
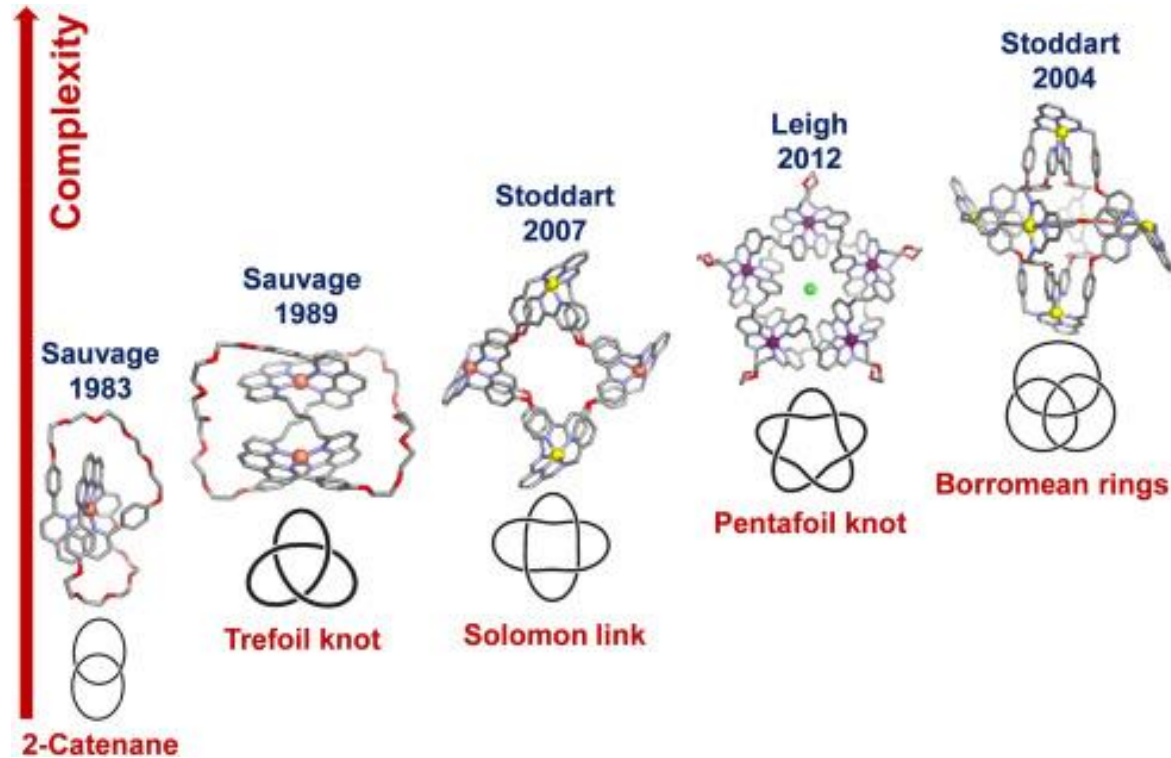
Crystal structures of links and knots ranging in their complexity from the first obtained catenane to the Borromean rings (Bilbeisi et al., 2017).

- 1980s – successful syntheses of first small-molecule knots

- Liu et al., 1981 – first knots found in natural molecules

- Knots form spontaneously in polymer chains of sufficient length and flexibility.

- Knots are observed in ~ 1% of proteins.

- Knots in proteins: stabilizing effect, functional sites, strongly conserved during evolution

- Occasionally tied in molecular models generated computationally; totally random.

- Any knots found in natural RNA molecules?

Can an RNA strand tie itself into a knot?

M. VanLoock, B. Harris, S. Harvey (**1998**) To Knot or Not to Knot? Examination of 16s Ribosomal RNA Models, *J Biomol Struct Dyn* 16(3): 709–713

- Topological knots in large native RNA structures are highly unlikely.

- Native RNA structures probably lack topological knots.

- Four existing models of the 30S ribosomal subunit were checked for knots in 16S rRNA (via Alexander polynomial).

- Three models had at least one knotted domain.

Conclusion: Models of large RNAs should be examined for knotting before publication.

Can an RNA strand tie itself into a knot?

C. Micheletti, M. Di Stefano, H. Orland (**2015**) Absence of Knots in Known RNA Structures, *PNAS* 112(7): 2052–2057

- ~6,000 RNA chains from PDB analysed for knots

- Three rRNAs solved by cryo-EM (3J5L, 4BW0, 4V9D) had knots.

- Knots found were suspected the artifacts as their higher-resolution homologs were unknotted.

Conclusion: Thermodynamic or kinetic mechanisms are likely at play for minimizing the entanglement of biologically viable structured RNAs.

Can an RNA strand tie itself into a knot?

A.S. Burton, M. Di Stefano, N. Lehman, H. Orland, C. Micheletti (**2016**) The Elusive Quest for RNA Knots, *RNA Biol* 13(2): 134–139

- Absence of knots in available RNAs discussed

- Why no knots have been found in RNA so far, despite the expectation that they should exist in Nature?

Conclusion: Some RNA sequences have the potential for forming knotted structures, as indicated by the properties of their predicted secondary structure.

Can an RNA strand tie itself into a knot?

W. Niemyska, S. Mukherjee, B. Gren, S. Niewieczerzal, J.M. Bujnicki, J.I. Sulkowska (**2024**) Discovery of a Trefoil Knot in the RydC RNA: Challenging Previous Notions of RNA Topology, *J Mol Biol* 436(6):168455
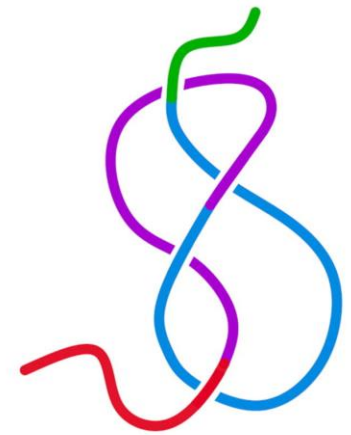
- Topological knot (trefoil) was found in experimental structure of the RydC RNA, a member of a family of short bacterial RNAs.

- It is the first topological knot found in the high-resolution structure of RNA.

- Postulate: this trefoil knot is a conserved feature of all RydC-related RNAs.
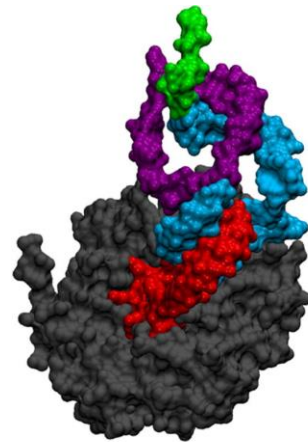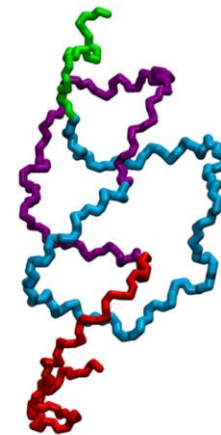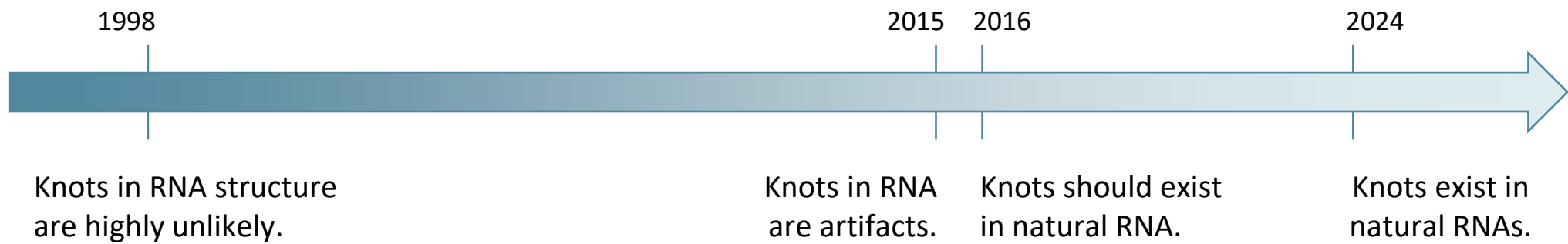

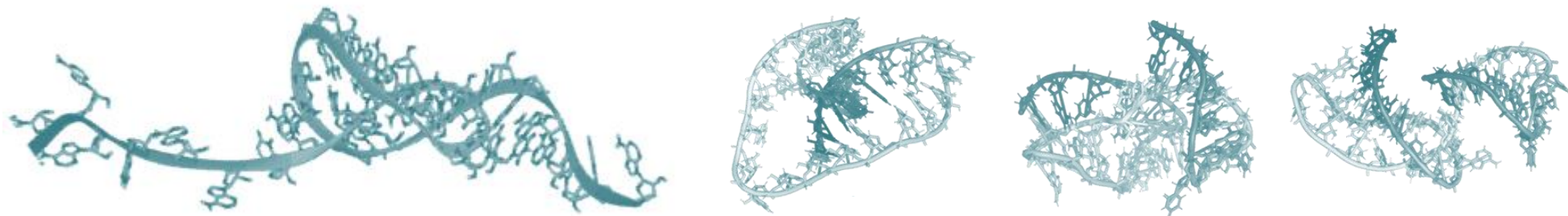Hfq/RydC complex   RydC RNA backbone   Knot formed by RydC

Topological knots in RNA structures
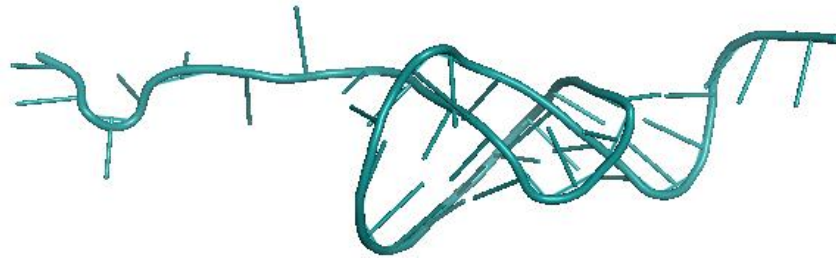
- In the study of natural RNAs

1998 — Knots in RNA structure are highly unlikely.

2015 — Knots in RNA are artifacts.

2016 — Knots should exist in natural RNA.

2024 — Knots exist in natural RNAs.

- Meanwhile in the field of RNA 3D structure prediction…

Topological knots in RNA 3D structure prediction
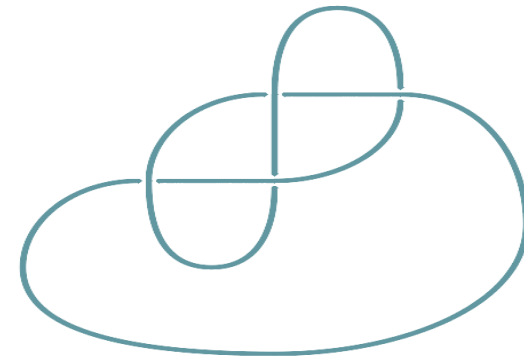
in silico generated model

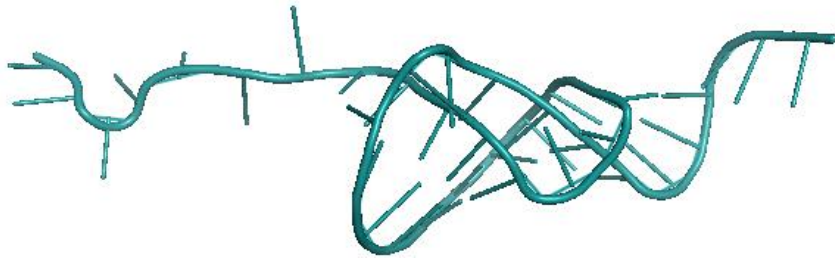rope

Figure-eight knot (Listing's knot)

- Alexander-Briggs notation: 4_1

- crossing number = 4
  (3rd-smallest possible, after unknot and trefoil)

- In sailing and rock climbing it serves as stopper knot;
  it stops a rope from running out of retaining device.

- In RNA structure – a kinetic trap

Snappy diagram of **4_1 knot**
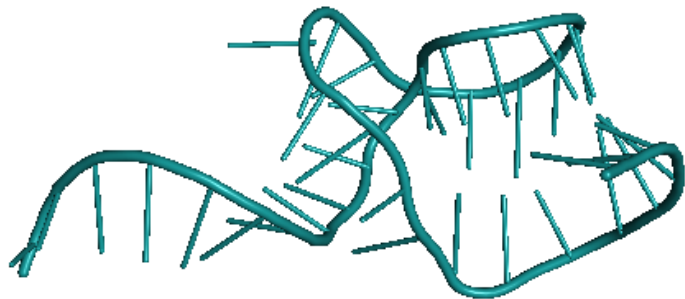
What happened here? A (pseudo)knot...
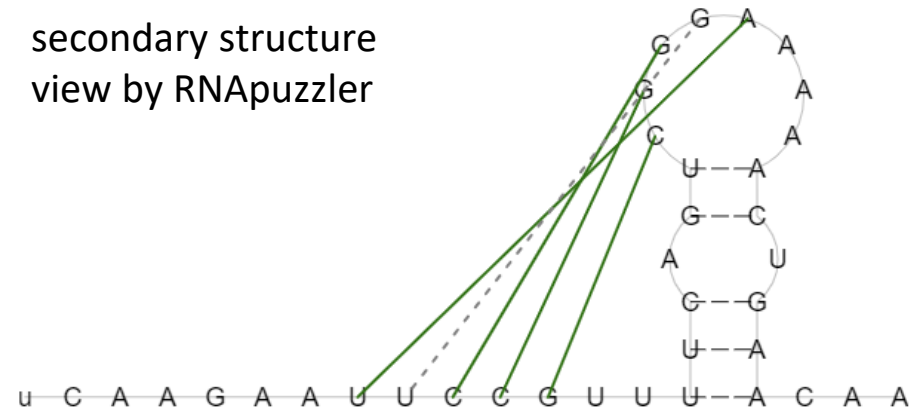
in silico generated model (2013)

input data

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

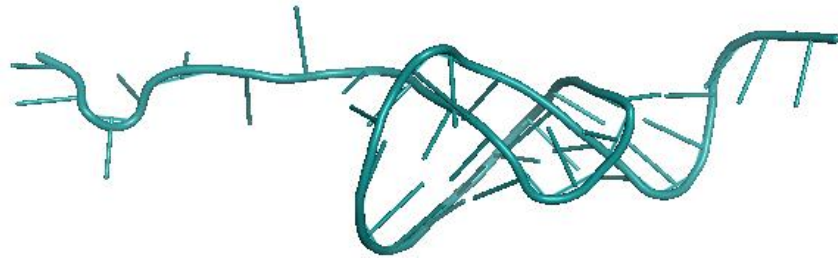*in silico* generated model (now)

secondary structure
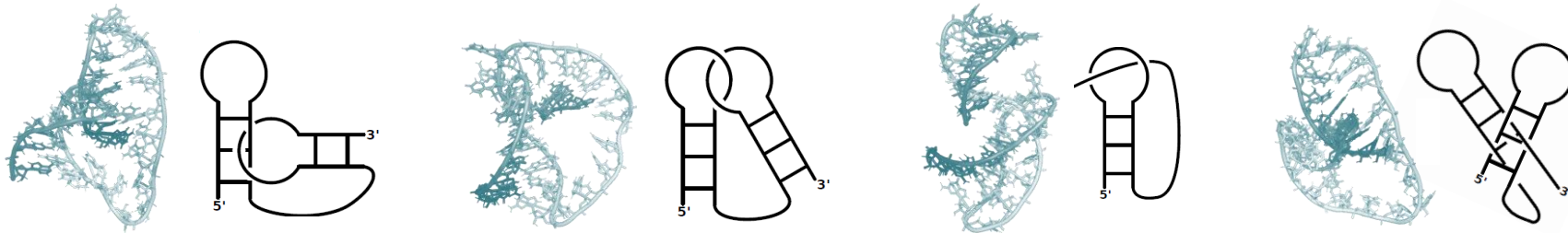view by RNApuzzler

Search for knotted elements (local perspective)

in silico generated model

← dangling end threaded through a loop

We took a closer look at RNA 3D models predicted in RNA-Puzzles...

Entanglements of structure elements

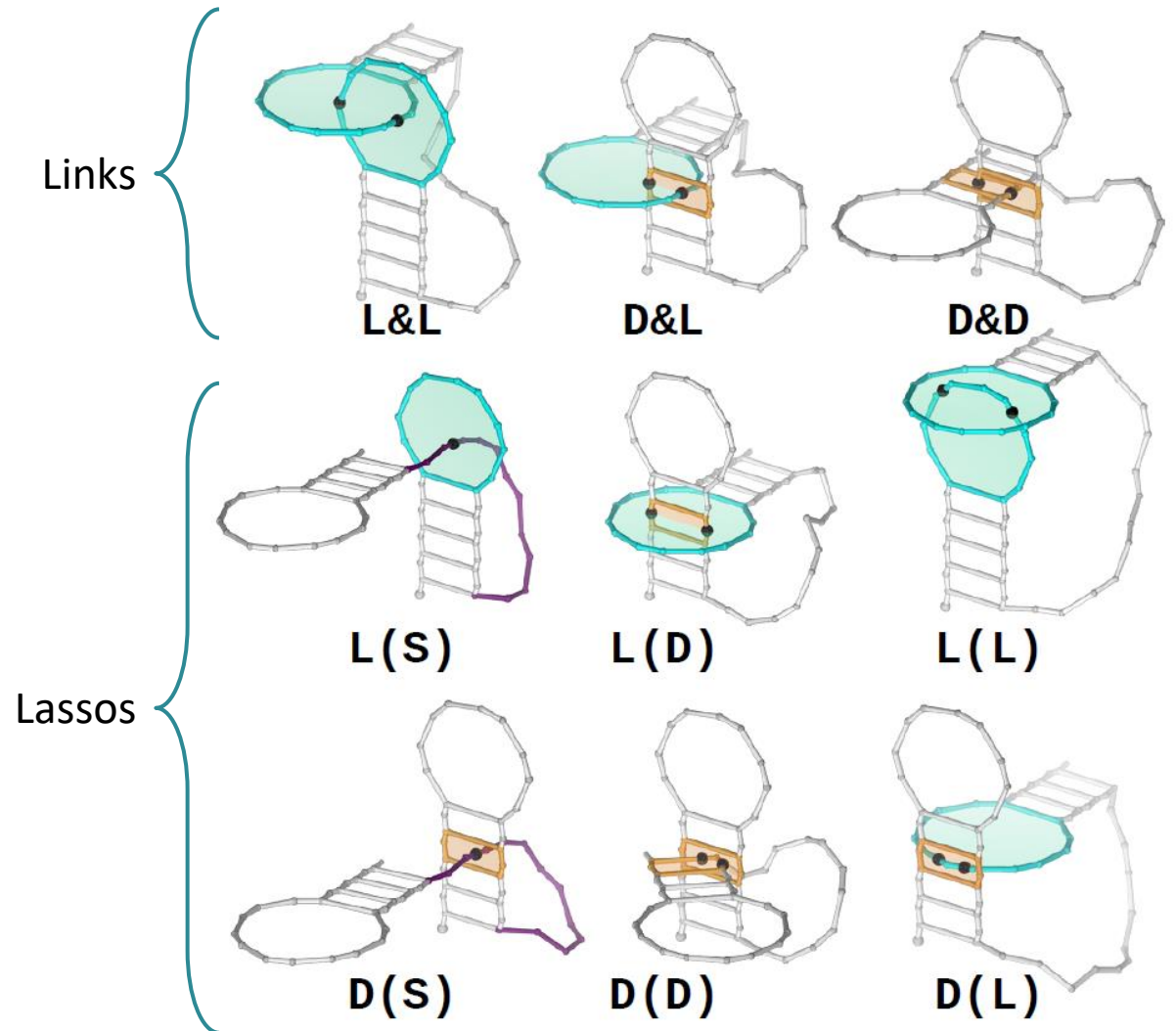RNA structure = assembly of structure elements

Entanglement = collision of two structure elements

Types of entanglements

- by topology
  - & link (interlace)
  - () lasso

- by colided structure elements
  - L loop
  - D dinucleotide step
  - S single strand

Each entanglement involves at least one closed element;
L, D – closed elements, S – open element.

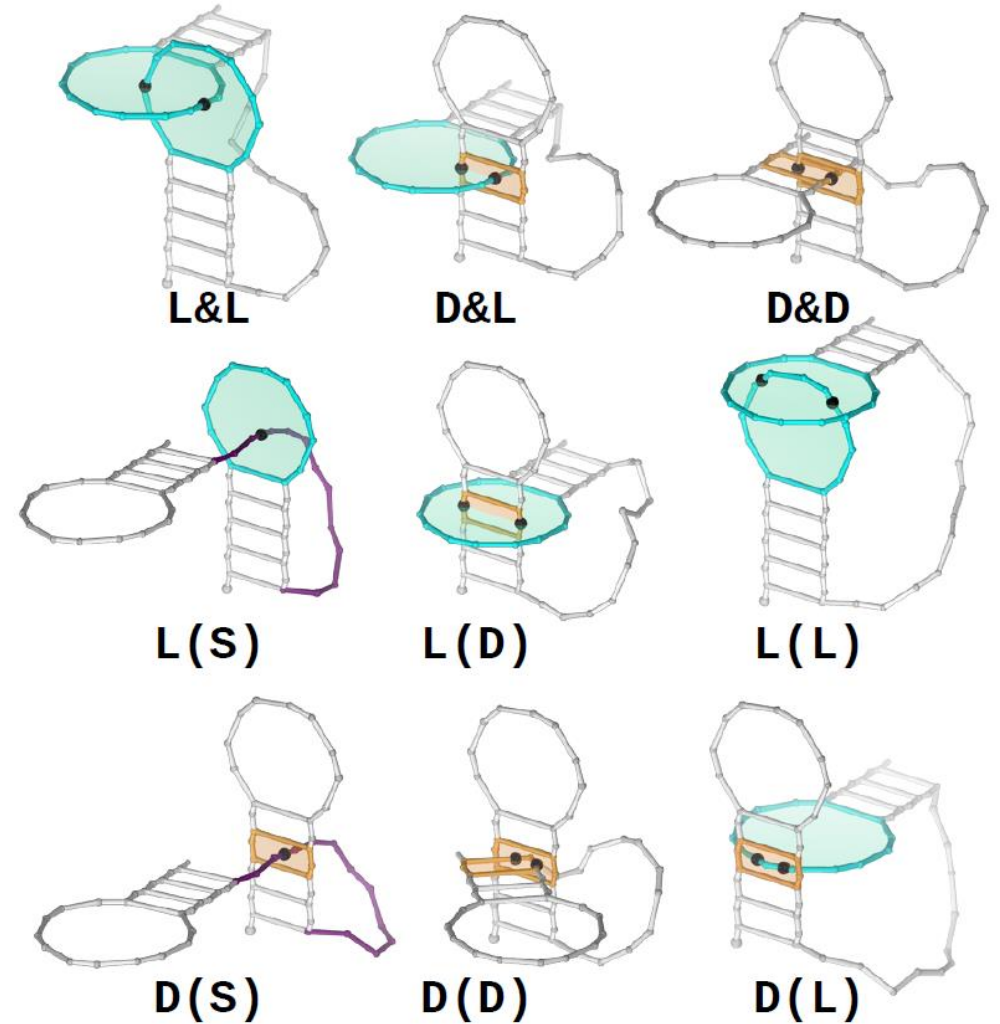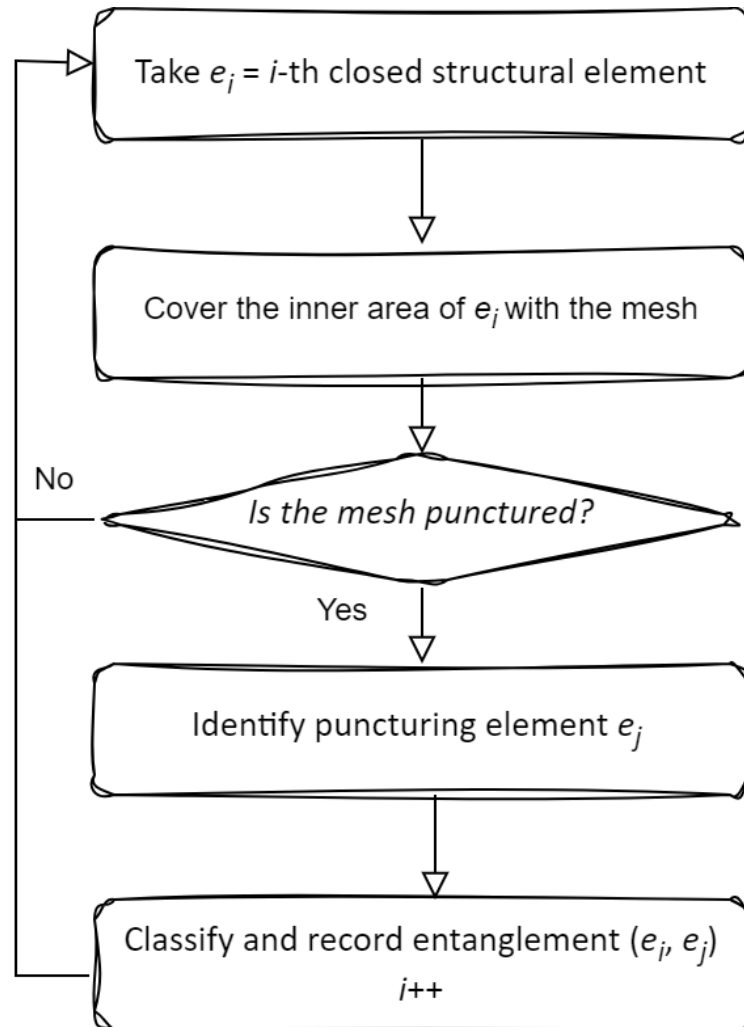Popenda *et al.* (2021) *Nucleic Acids Research* 49(17):9625–9632

Links

L&L   D&L   D&D

Lassos

L(S)   L(D)   L(L)

D(S)   D(D)   D(L)

Entanglements of structure elements

Which of these are possible?

Which are misfolds?

- Links
- Shallow lassoes (depth $\leq$ 5 nts)
- Deep lassoes (depth > 5 nts)

L&L   D&L   D&D

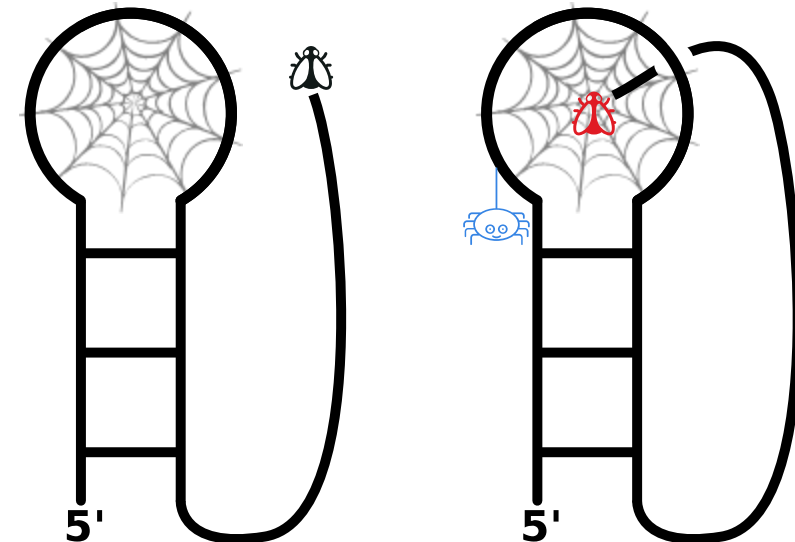L(S)   L(D)   L(L)

D(S)   D(D)   D(L)

Algorithm to detect entanglements – an idea

Closed structure elements:

- loops (apical loops, n-way junctions, bulges, internal loops)
- dinucleotide steps

Algorithm to detect entanglements at work

movie showing the algorithm's performance…

Mariusz Popenda

RNAspider web server: https://rnaspider.cs.put.poznan.pl

Luwanski *et al.* (2022) RNAspider: a webserver to analyze entanglements in RNA 3D structures, *Nucleic Acids Research* 50(W1):W663-W669

Search for entanglements in RNA 3D structures

RNA-Puzzles

- 22 target RNA structures (X-ray);
  all without entanglements
- **1,028** predicted RNA 3D models

CASP15

- 13 target RNA structure;
  all without entanglements
- **1,646** predicted RNA 3D models
  (structured)

RCSB PDB PROTEIN DATA BANK

- **5,348** experimental
  RNA structures

https://github.com/RNA-Puzzles

https://predictioncenter.org/casp15

https://www.rcsb.org/

Data for the first experiment collected in December 2022.

Entanglements in predicted & experimental RNA 3D structures

Percentage of structures that are ☐ non-entangled ☐ entangled

RNA-Puzzles (1028) * ^ : 14% entangled, 86% non-entangled
CASP15 (1646) * : 8% entangled, 92% non-entangled
PDB (5348) ^ : 18% entangled, 82% non-entangled

* Target structures not entangled   ^ Data collected in December 2022

Record breakers (number of entanglements per structure)
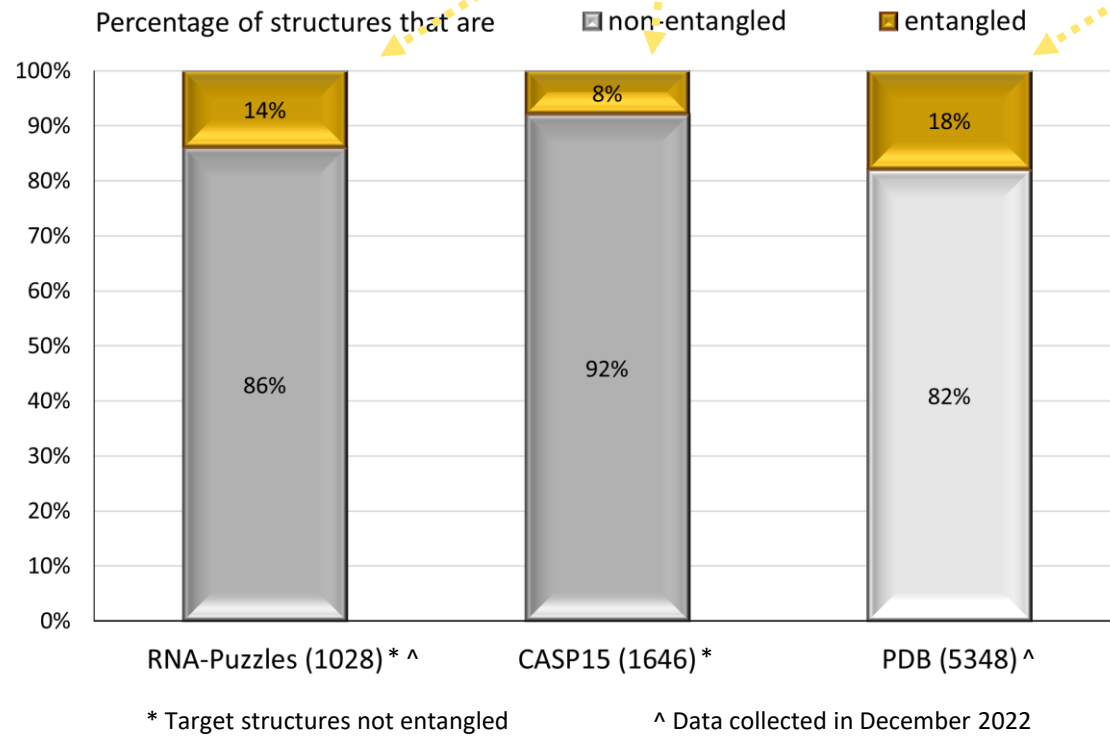
- 70 entanglements in R1138TS235_2 model
  (predicted in CASP15 by a nonML method)

- 21 entanglements in 1C2W structure
  (experimental structure, CryoEM, resolution: 7.5 Å)

Distribution of entanglements by type

RNA-Puzzles: lasso 82%, interlace 18%

CASP15: lasso 63%, interlace 37%

PDB: lasso 99,5%, interlace 0,5%

☐ lasso
☐ interlace

Interlaces in PDB - only in *low-resolution* structures.

Classes <u>not represented</u> in the analysed datasets:

- RNA-Puzzles: D(S), D(L)
- CASP15, PDB: −
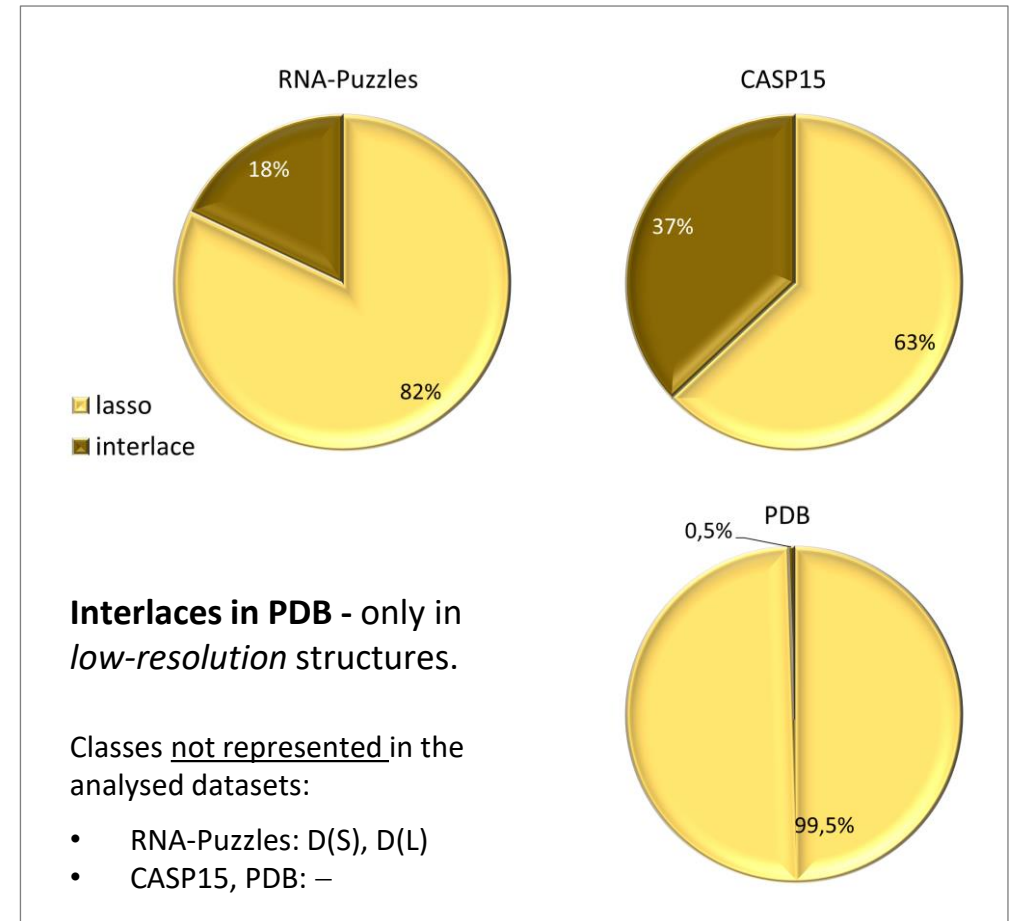
Occurrence of entanglements in predicted structures

Pseudoknot-free
469

447

Entangle-free
907

22

460

Entangled
143

121
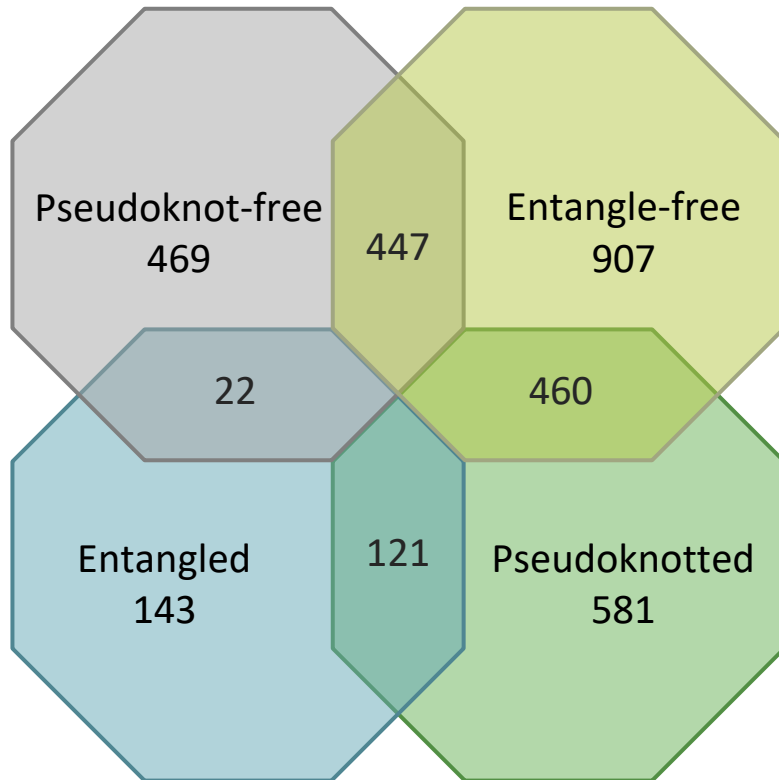
Pseudoknotted
581

❑ **No correlation** with RMSD, Clash score, molecule size

❑ **Strong correlation** with pseudoknots

73% of all lassos

60% of all links

formed in the **pseudoknotted** structures

The probability of entanglement in pseudoknotted RNA is **4x higher** than in non-pseudoknotted structure.

Maciej Antczak

Tomasz Zok

Mariusz Popenda

Joanna Sarzynska

Knotted column (flat knot/ Hercules knot)
Trento Cathedral (Italy)

Entangled structure elements ✅

Topological knots ... ❓

30.01-02.02.2023, Structure
and topology of RNA in living
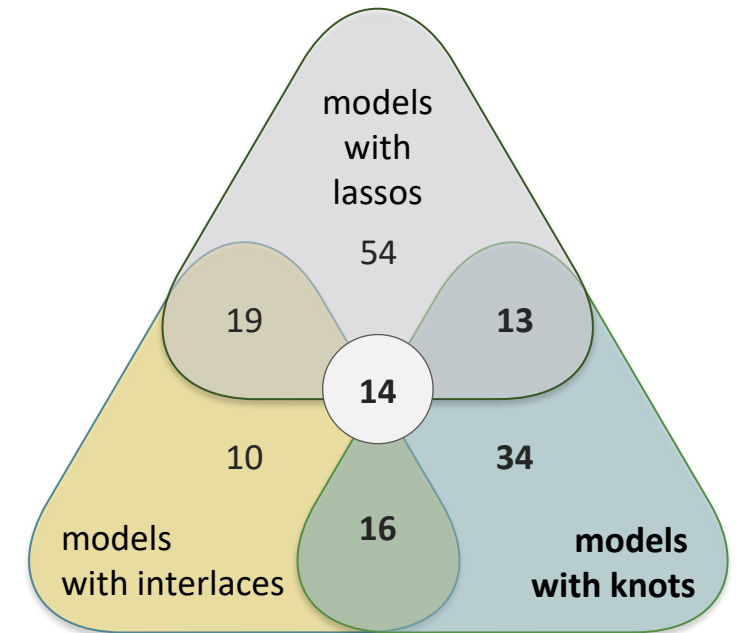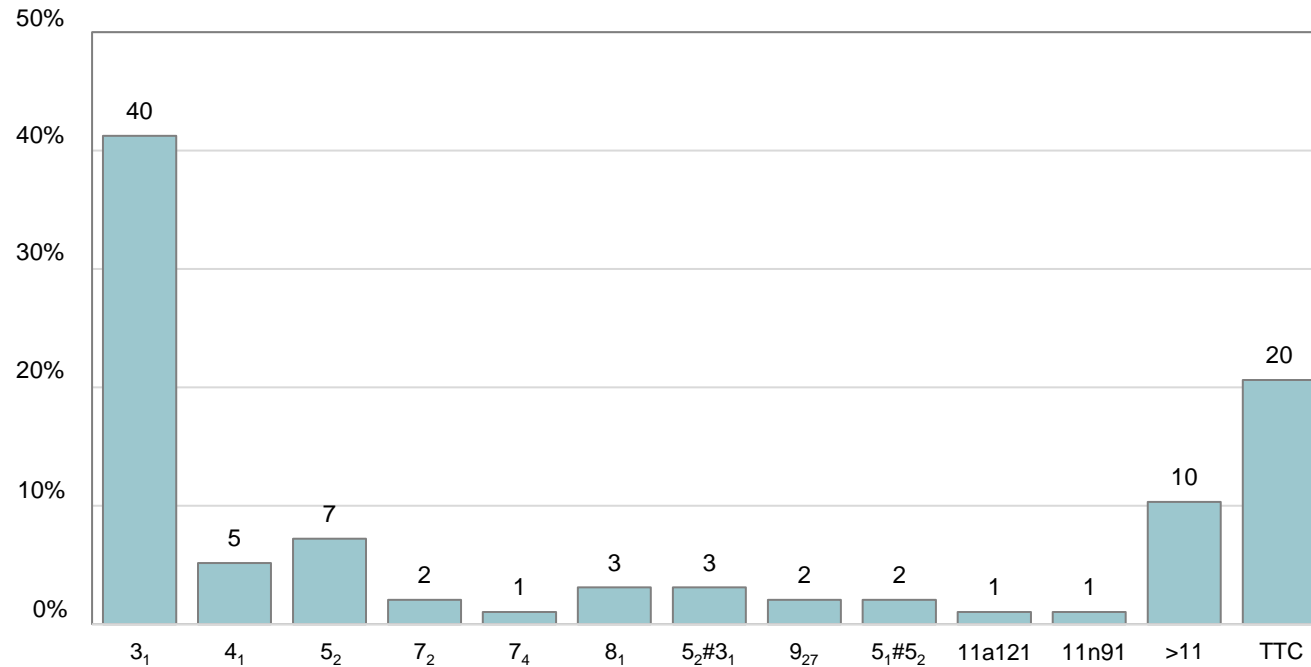systems, Trento, Italy

Wanda Niemyska

Joanna I. Sułkowska

Bartosz A. Gren

Topological knots in CASP15 predictions

77 predictions with topological knots
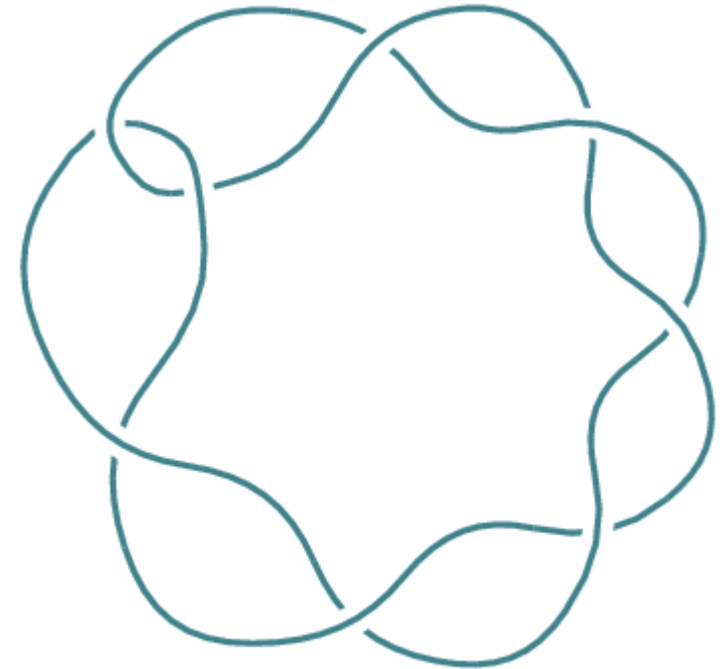in the set of 1646 models submitted to CASP15

Topological knots in CASP15 predictions

Target: non-natural, 374 nt

Misfolded or not?

Model R1136TS185_2
by human group using ML
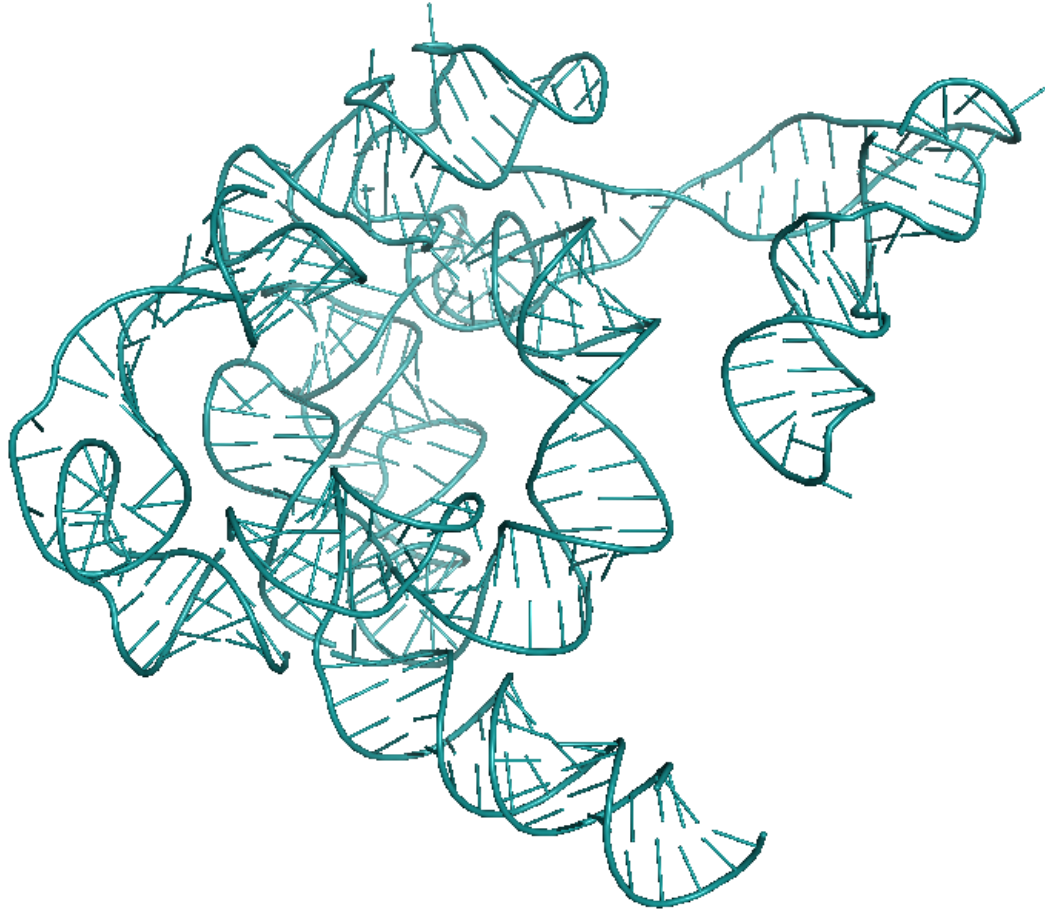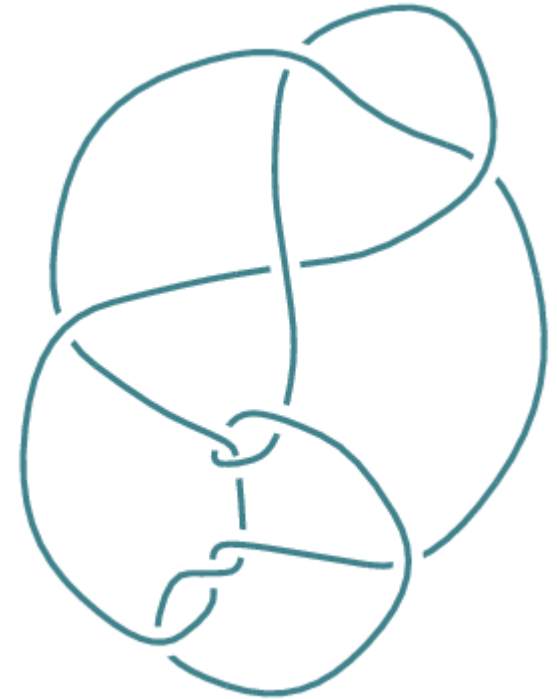
8_1 knot

Topological knots in CASP15 predictions

Target: non-natural, 374 nt (same target, another method)
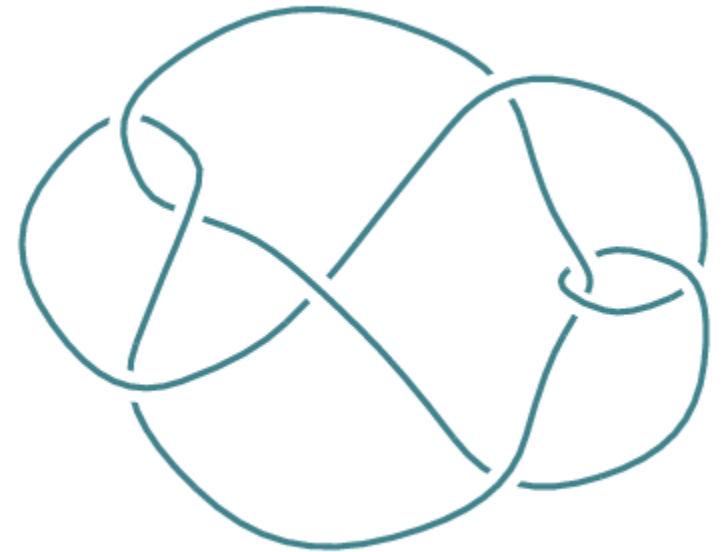

Model R1136TS054_3
by human group using ML

11n91 knot

Topological knots in CASP15 predictions
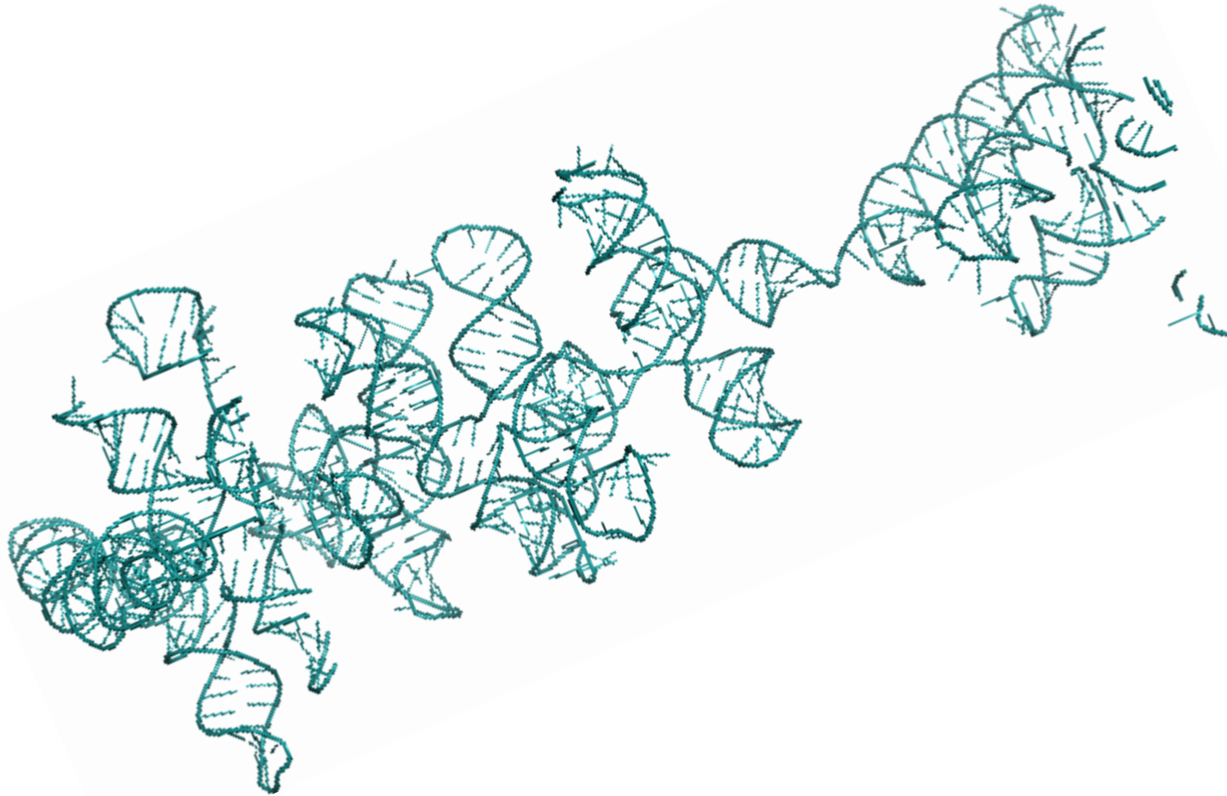
Target: non-natural, 363 nt

Model R1126TS444_1
by human group using nonML method

9_{27} knot

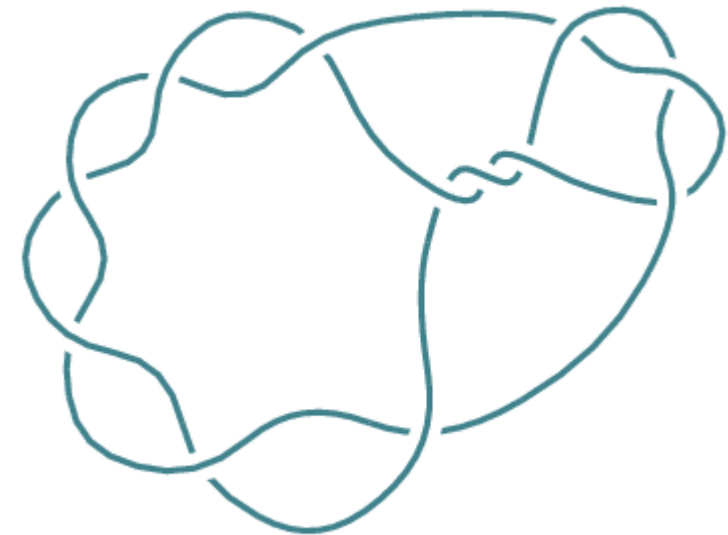# Topological knots in CASP15 predictions

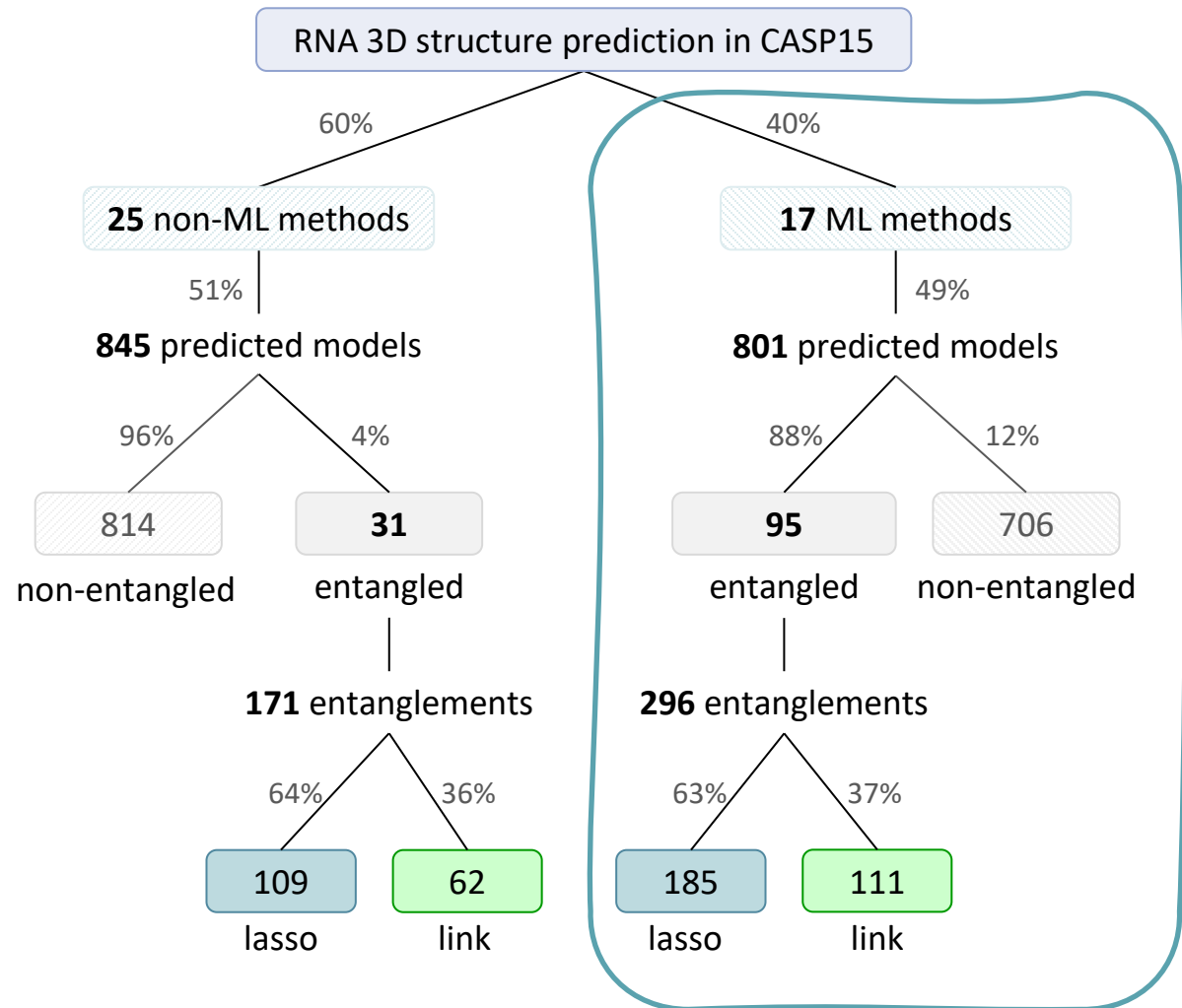Target: non-natural, 720 nt

Model R1138TS235_1
by human group using nonML method

12a1282 knot

CASP15: method-focused analysis

ML methods (= 40% of all methods) generated 63% of entanglements, including complex knots!

Analysis done using:

- RNAspider (Luwanski et al., *Nucleic Acids Res* 2022)

- Topoly (Dabrowski-Tumanski et al., *Brief Bioinform* 2021)
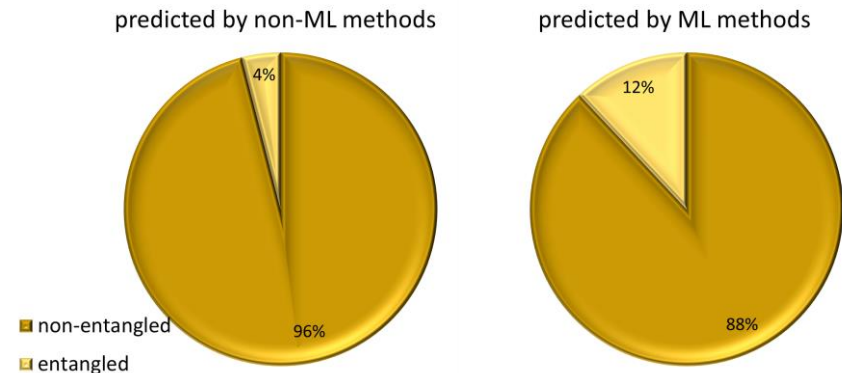
CASP15: target-focused analysis

Distribution of entangled 3D RNA structure predictions by target. Target structures are grouped by difficulty.

(A) Entanglements of structure elements

(B) Topological knots

Targets

- Easy
- Medium
- Difficult
- Non-natural

(A) Links and lassos | Trivial topology

(B) Knotted | Unknotted
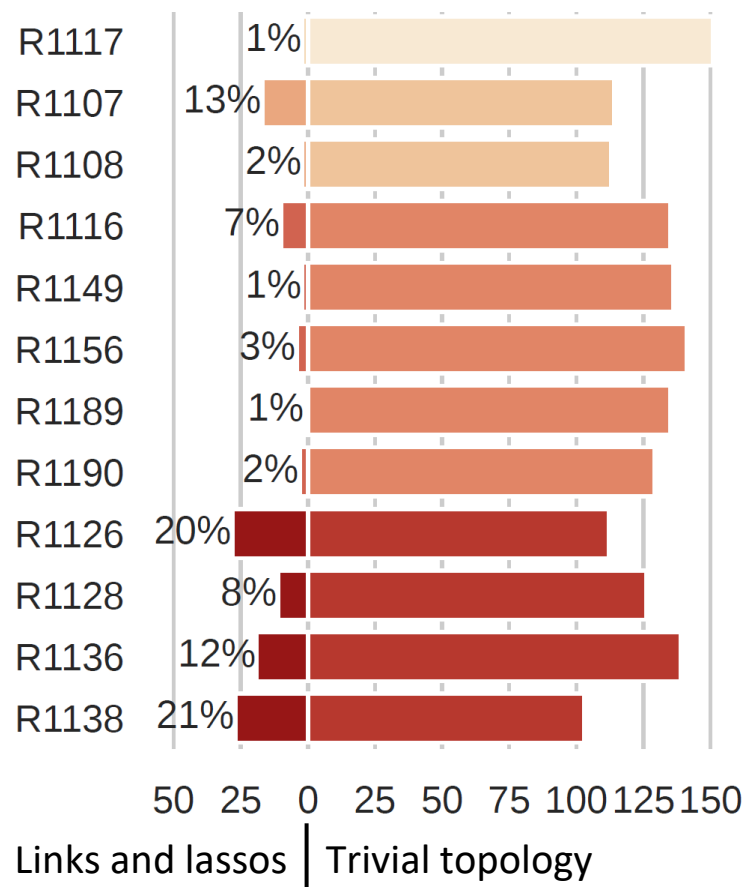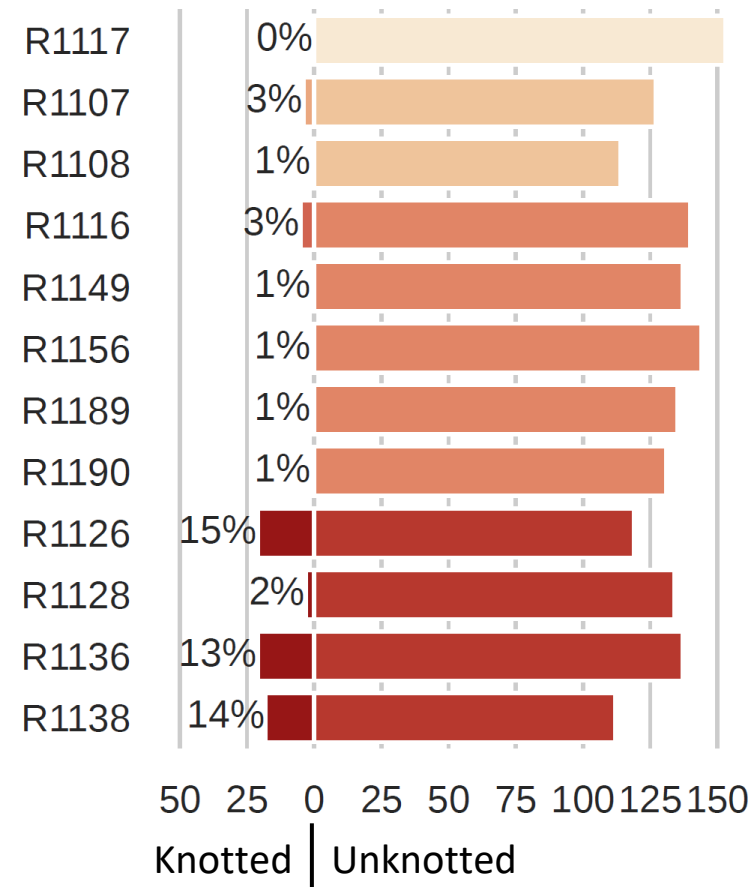
In CASP16...

Example predictions by AlphaFold 3
for R1252 (EM, 520 nt)

8 × link: 5 x D&D, 2 x L&L, D&L
2 × lasso: D(D), D(L)

6 x link: 4 x L&L, 2 x D&L
7 x lasso: 5 x L(L), 2 x L(D)

In CASP16...

Example predictions by AlphaFold 3
for R0250 (EM, 744 nt)
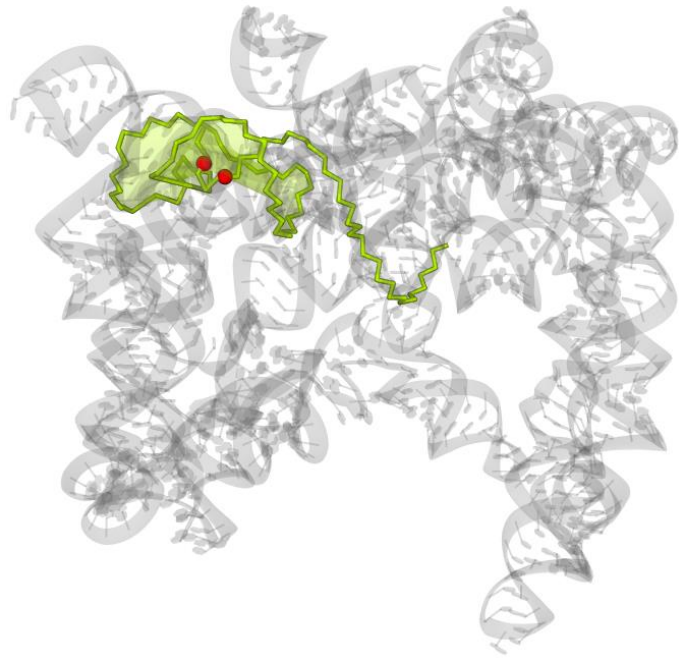
1 x lasso: L(S.)
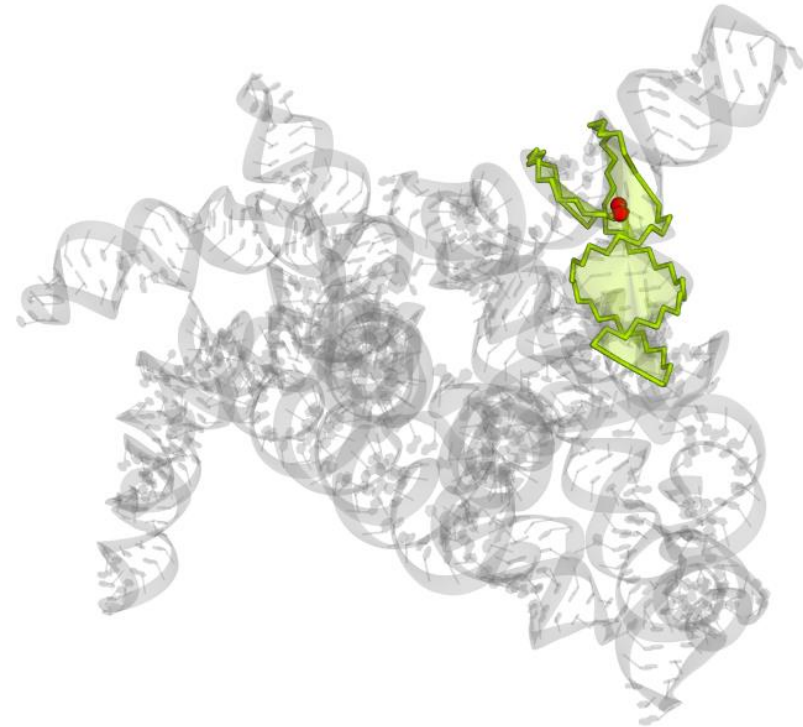
1 x lasso: L(L)

Remedy for unwanted entanglements

Energy minimization

CASP15: Target R1138 (720 nt)

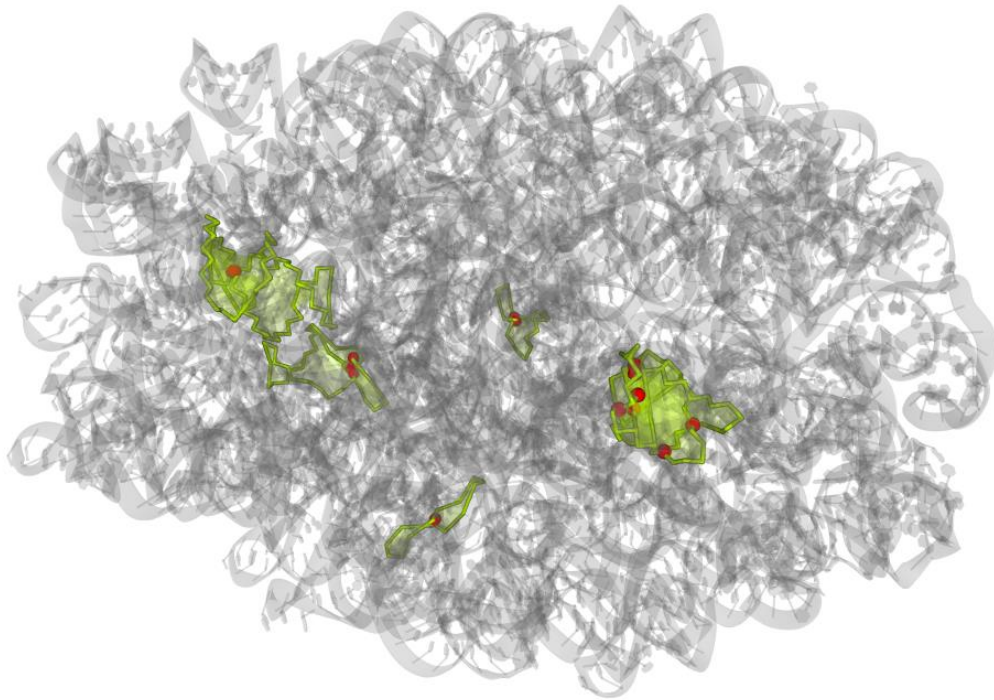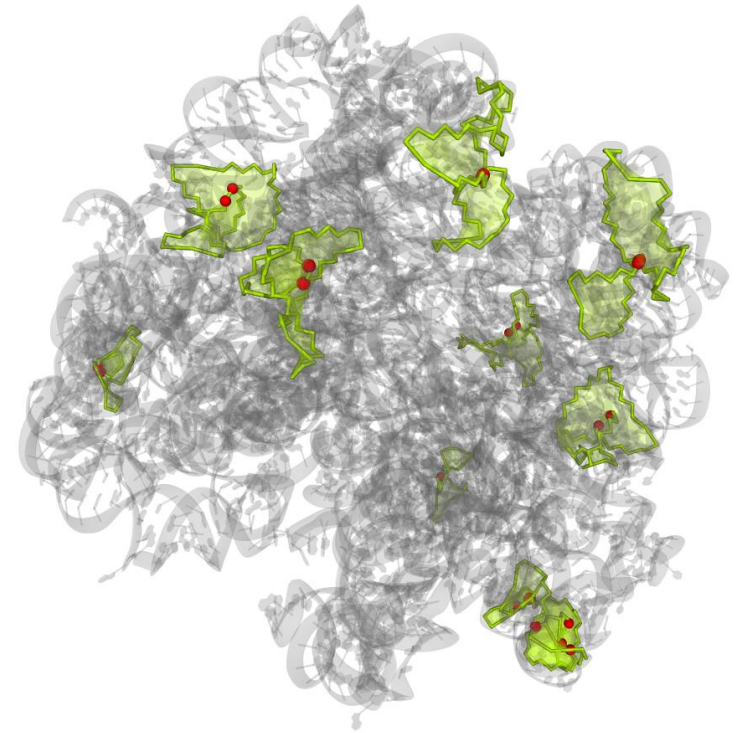Model R1138TS081_4
(RNApolis group)
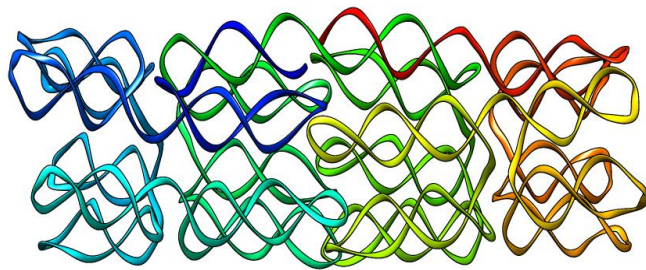
Model evaluation before and after energy minimization (better values in shaded cells)

	# entanglements	INF$_{all}$
Initial model	55	0.912
Refined model	0	0.984

Number of entanglements in consecutive iterations of energy minimization by **CYANA** (minimization in torsion angle space):
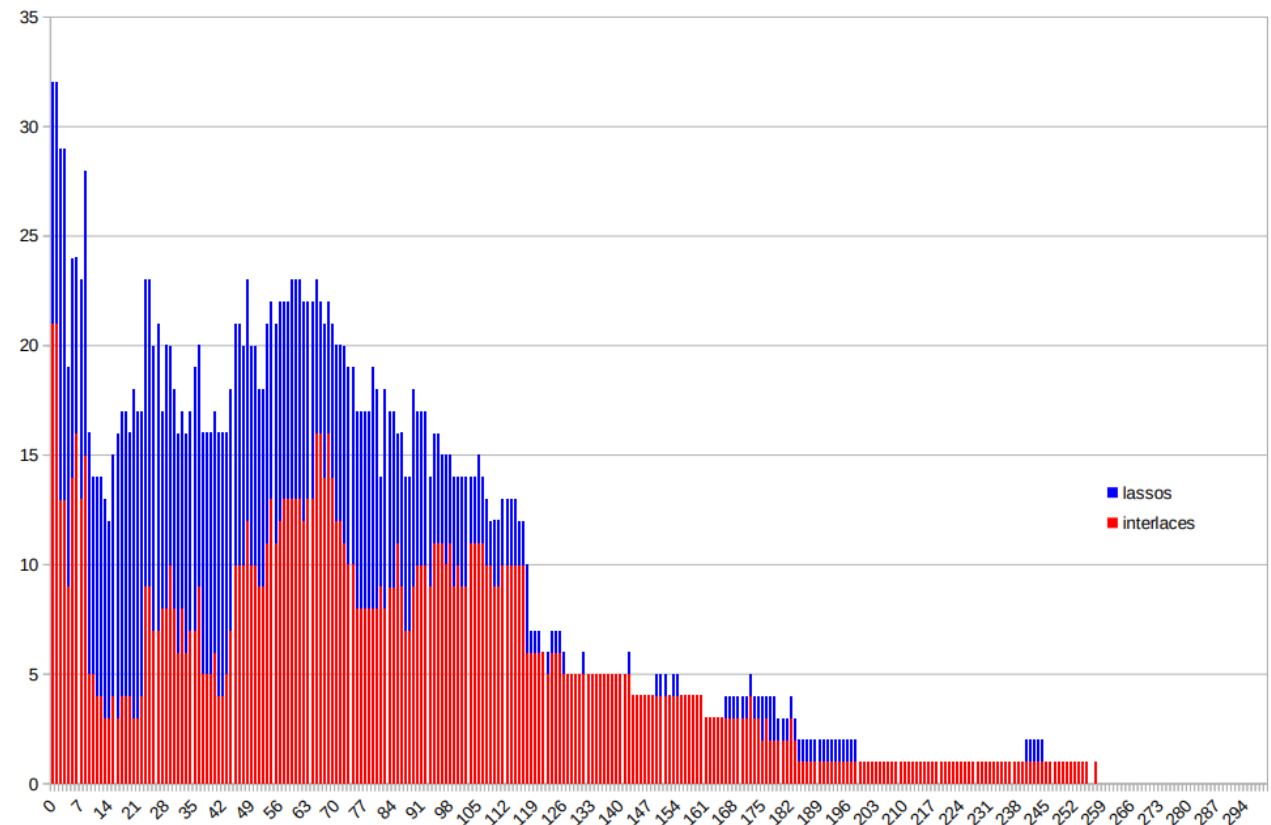
Remedy for unwanted entanglements

CASP15: Target R1116 (157 nt)

Model R1116TS470_1
(CoMMiT-human group)

This model contains
7 entanglements:

- 4 x link: D&D
- 3 x lasso: D(D)

Refinement (with disentangling):

- **SPQR** - SPlit and conQueR for RNA structure prediction *by Simon Poblete*
- **RNAtuner** *by Maciej Antczak*
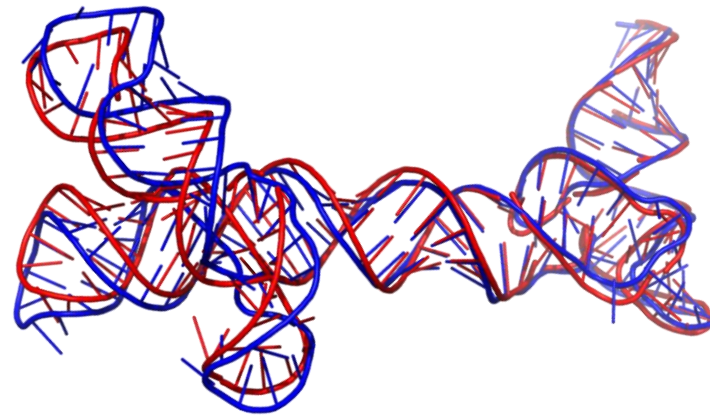
No entanglements after refinement!

Model evaluation before and after SPQR
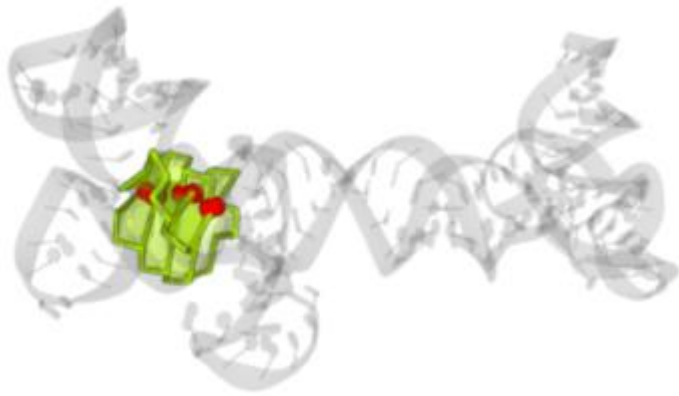refinement (better values in shaded cells) ⟶

	Initial model	Refined model
INF$_{WC}$	0.821	0.928
INF$_{NWC}$	0	0.105
INF$_{stack}$	0.699	0.615
INF$_{all}$	0.615	0.645
RMSD	17.373	17.088
DI	26.939	26.346
P-value	2.99E-04	1.65E-04
Clash score	123.74	17.07

Remedy for unwanted entanglements

SPQR-driven disentanglement
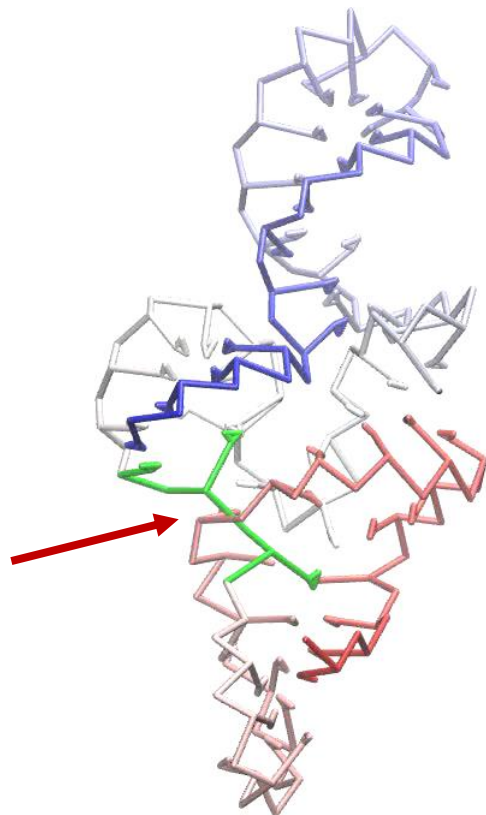
1. SPQR energy minimization, until no clashes and broken bonds are present.

2. A short simulation including repulsive virtual sites between the structure elements in collision.

3. SPQR energy minimization (in case that some clashes appeared in step 2).

4. A short steered simulation to pull back the positions and orientations of the nucleobases to resemble the original structure.

5. Coarse-grained to full atom model tranformation.

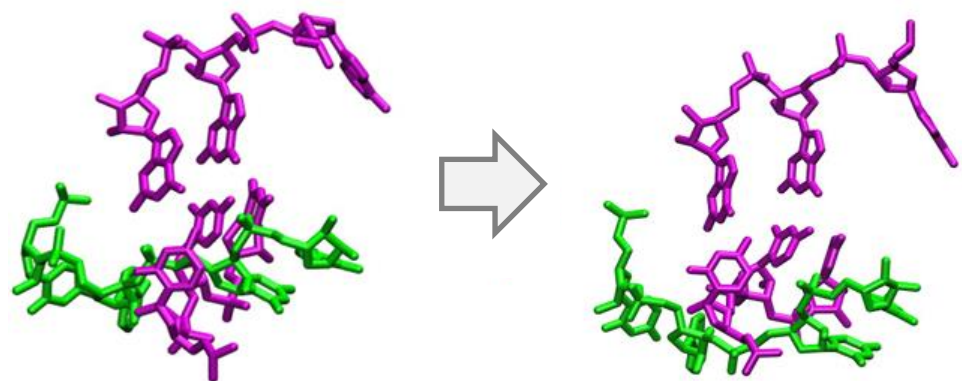6. Energy minimization in Gromacs (using the Amber force field and restraints on the positions).

Simon Poblete

Mikolaj Mlynarczyk

Poblete *et al*. (2024) Unknotting RNA: a method to resolve computational artifacts, *working paper*
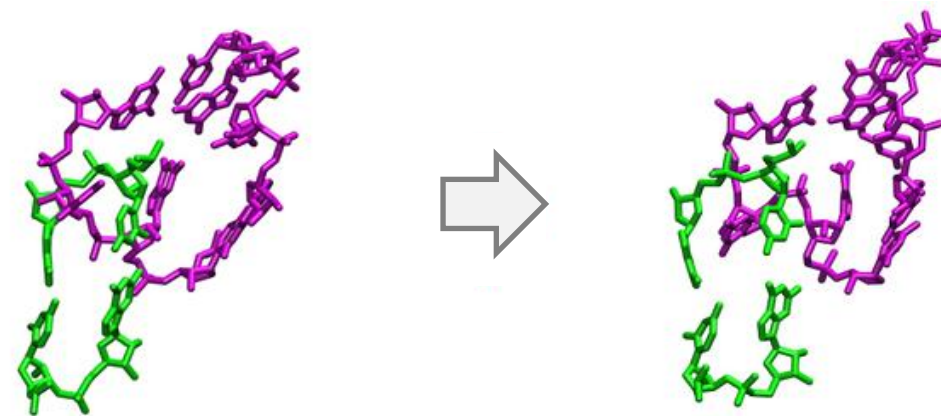
Remedy for unwanted entanglements

D(S)-type lasso removal

D&L-type link removal

First experiment with SPQR-driven disentanglement of CASP15 predictions

	D(S)	L(S)	L(D)	L(L)	D&D	D&L	L&L	Total
Total occurrences	12	38	40	13	12	13	17	145
Removed	12	17	10	3	12	12	11	77

Summary

o   Entanglements and topological knots are found in the experimental and predicted RNA 3D models.

o   Most predicted entanglements are modeling artifacts.

o   Many predicted topological knots are misfolds

o   ML-based predictors are more likely to entangle structures.

o   Automated identification of entanglements and knots

  •   RNAspider (entangled structure elements) https://rnaspider.cs.put.poznan.pl

  •   Topoly Python package (knots) https://github.com/ilbsm/CASP15_knotted_artifacts

  •   KymoKnot (knots) http://kymoknot.sissa.it/kymoknot/interactive.php

o   Automated removal of entanglements is possible

Acknowledgments

Permanent team

- Maciej Antczak
- Mariusz Popenda
- Joanna Sarzyńska
- Tomasz Zok
- Marta Szachniuk

RNAspider development

- Vladyslav Hlushchenko
- Kamil Luwanski
- Daniil Martsich

Topological knot analysis

- Bartosz Gren
- Joanna Ida Sułkowska

Disentanglement

- Simon Poblete
- Mikolaj Mlynarczyk