Vfold Modeling of RNA Targets in CASP16

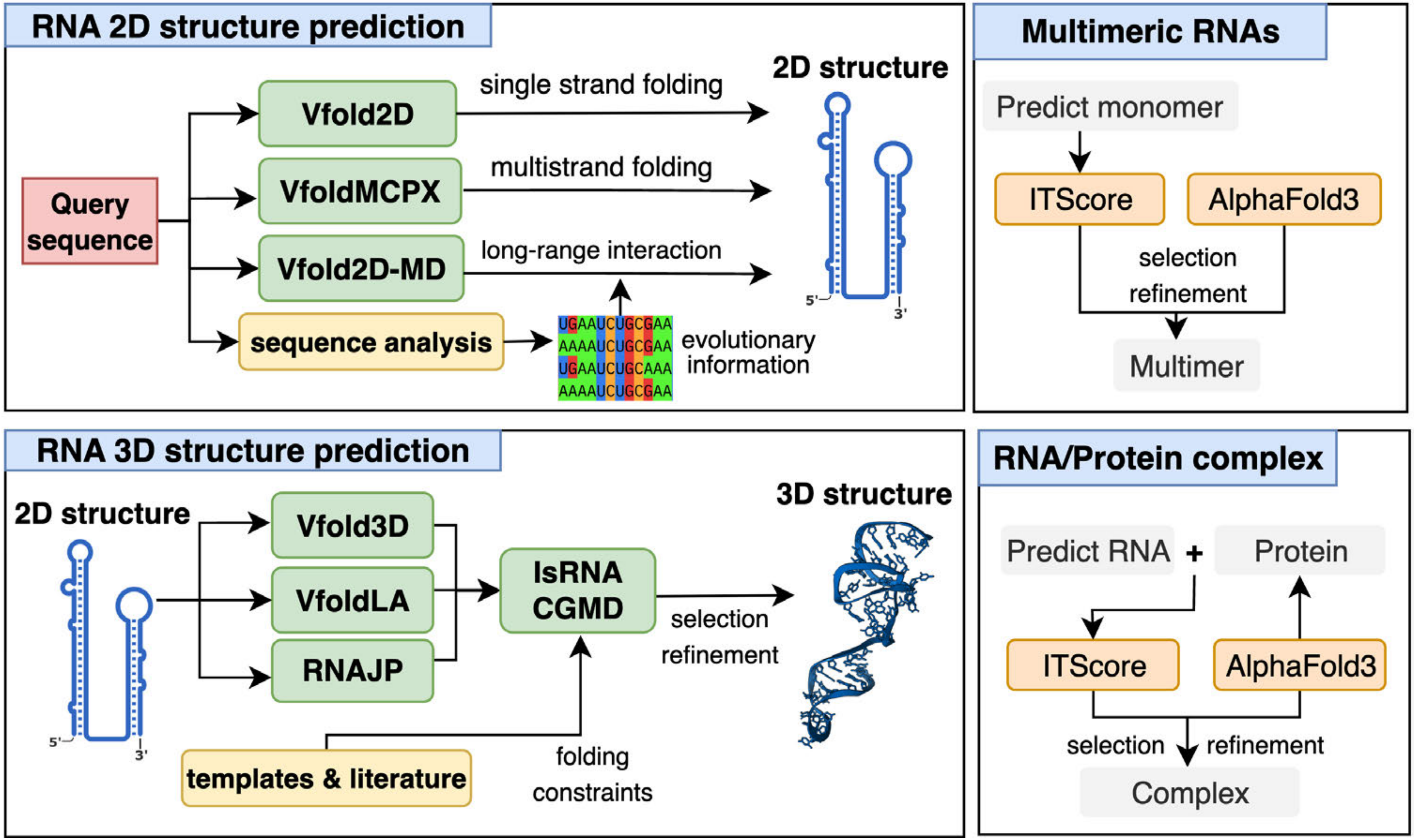
Shi-Jie Chen[1]

Sicheng Zhang[1], Jun Li[2], Yuanzhe Zhou[1]

[1] Department of Physics, Department of Biochemistry,
Institute for Data Science and Informatics
University of Missouri, USA
[2] Great Bay University, China

Mizzou
University of Missouri

AF3-assisted physics-based modeling

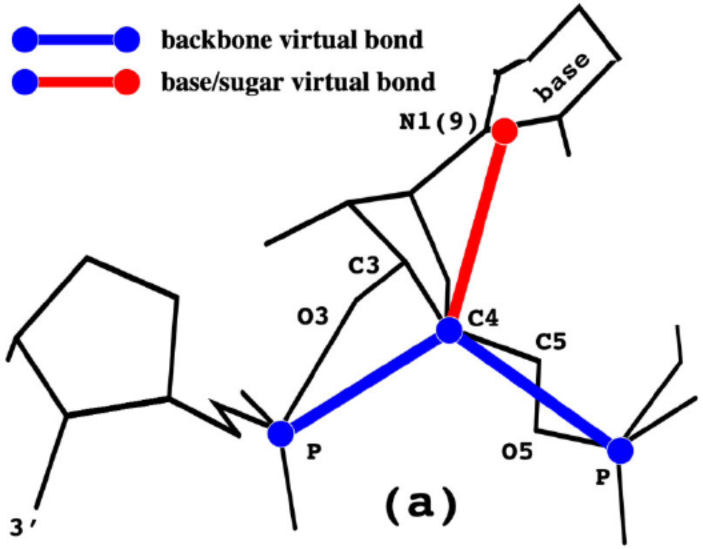


ITScore for docking: Zou lab

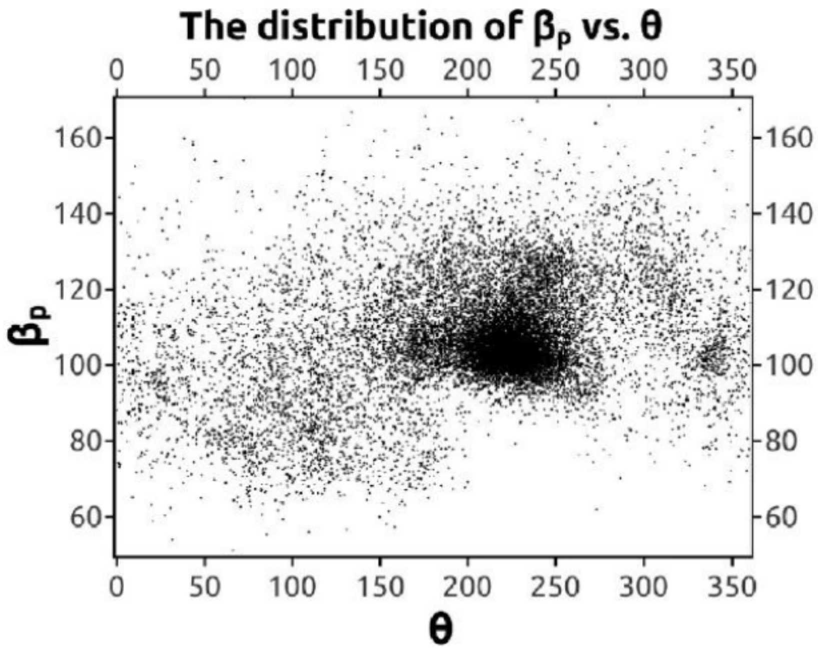
Vfold toolbox – Physics-based models

To predict 2D structure from the sequence

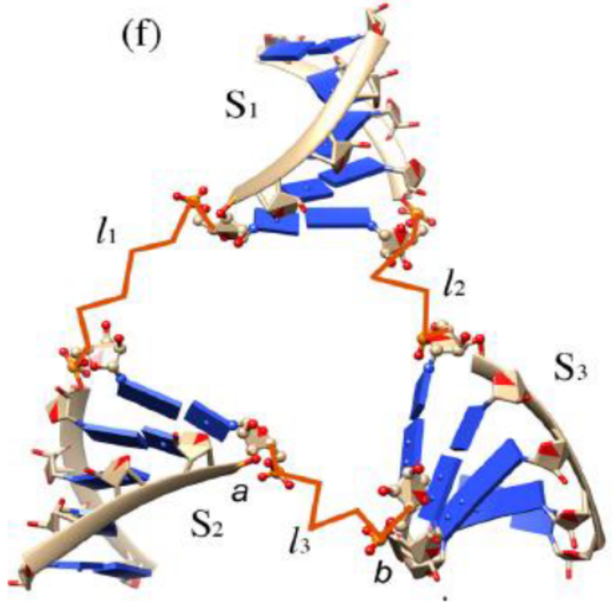
Free energy modeling: **Vfold2D** (monomer) **VfoldMCPX** (multimer)



A virtual bond (4 beads) representation of RNA conformations



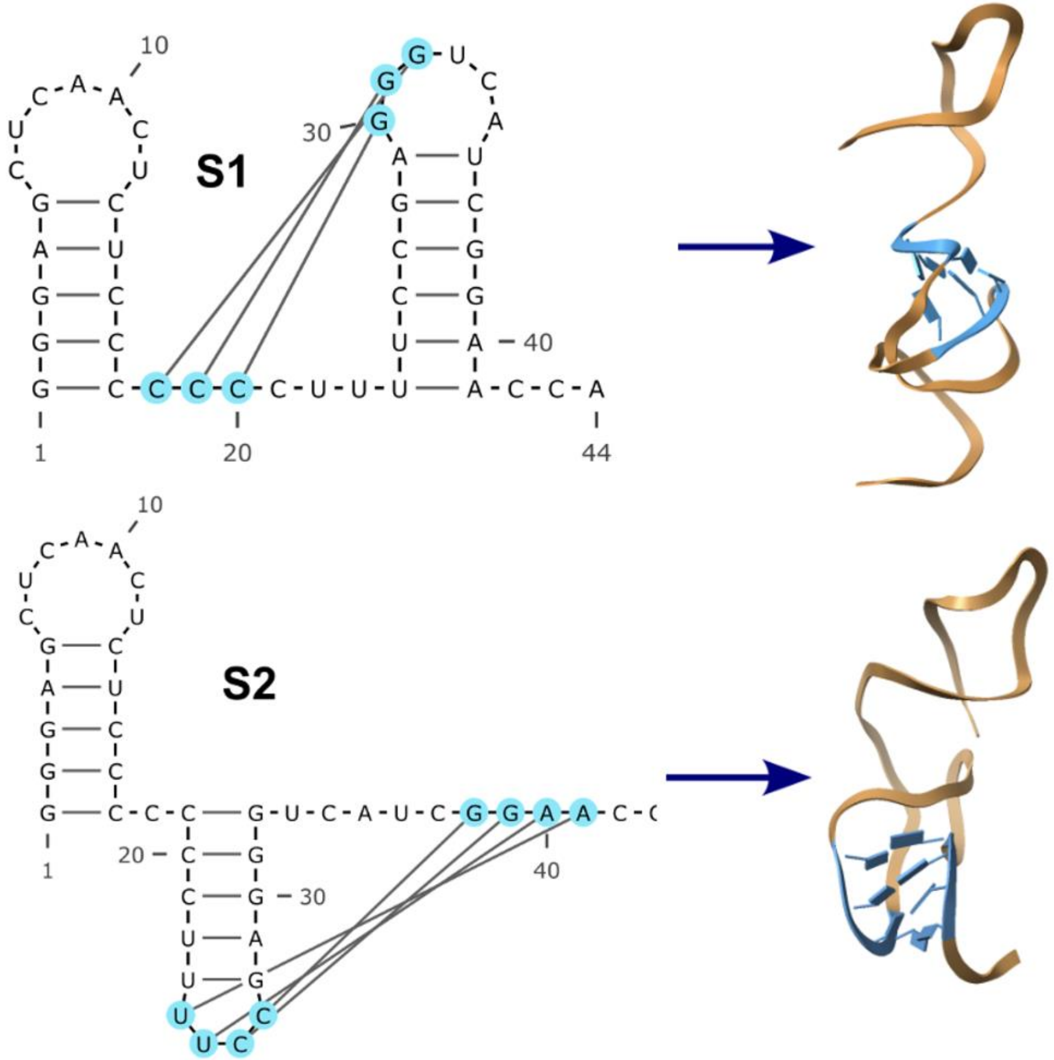
MC sampling guided by know structures



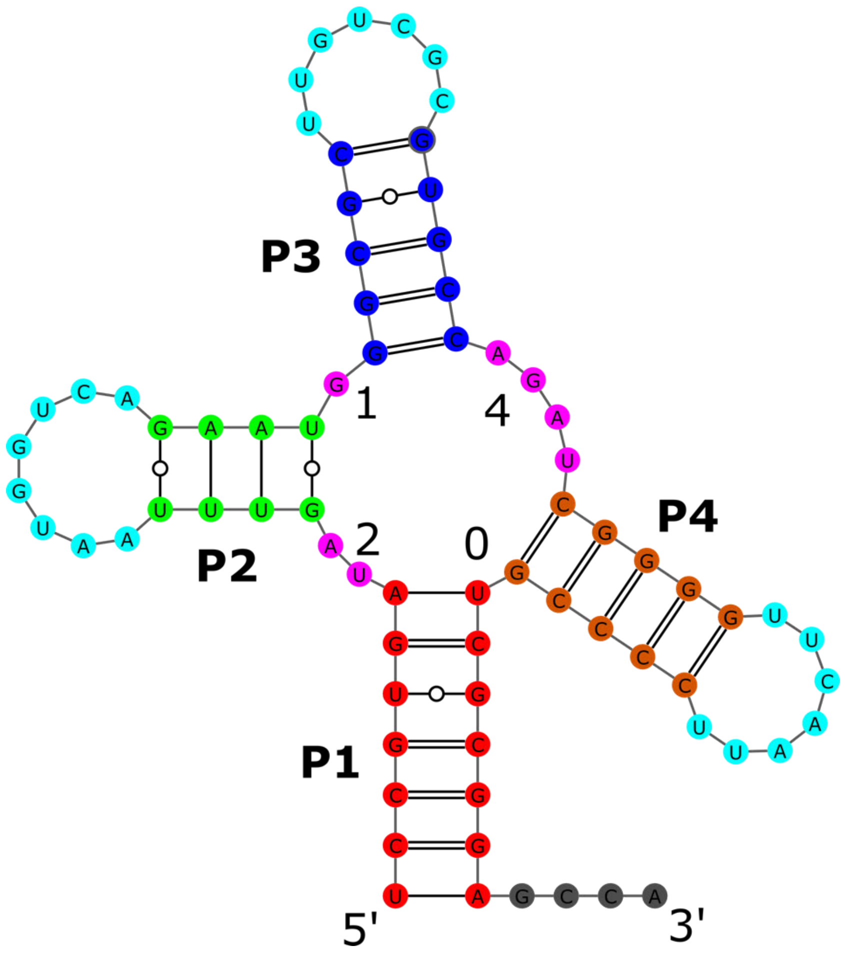
Free energy parameters for motifs

To predict long-range kissing interactions

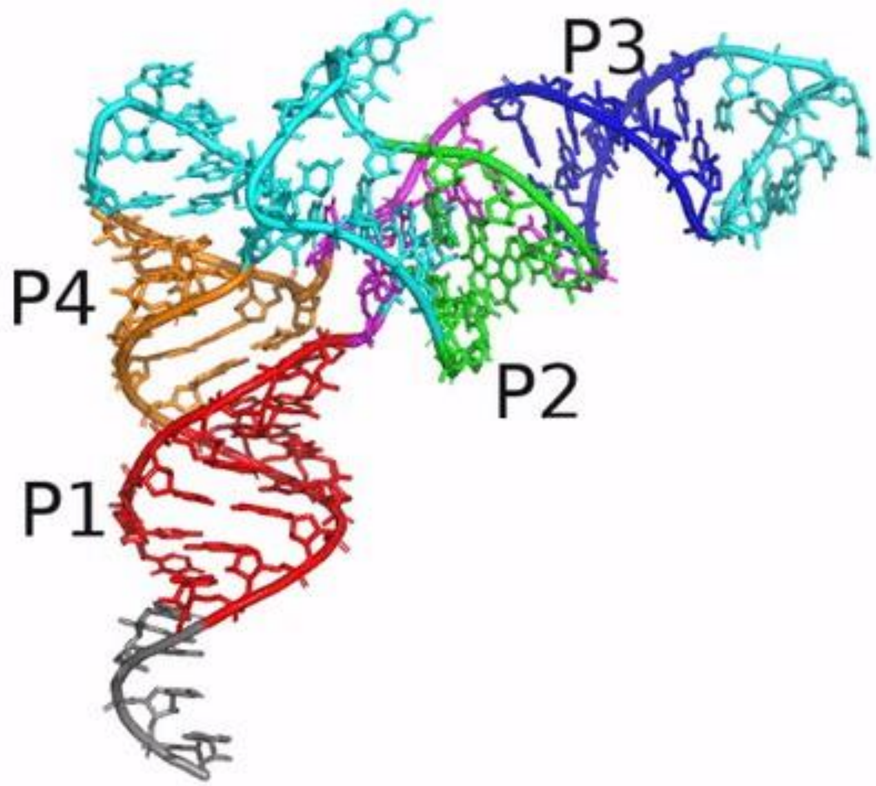
Vfold2D-MD (unpublished)



To generate (initial) 3D from 2D
Template-based modeling: Vfold3D & VfoldLA:

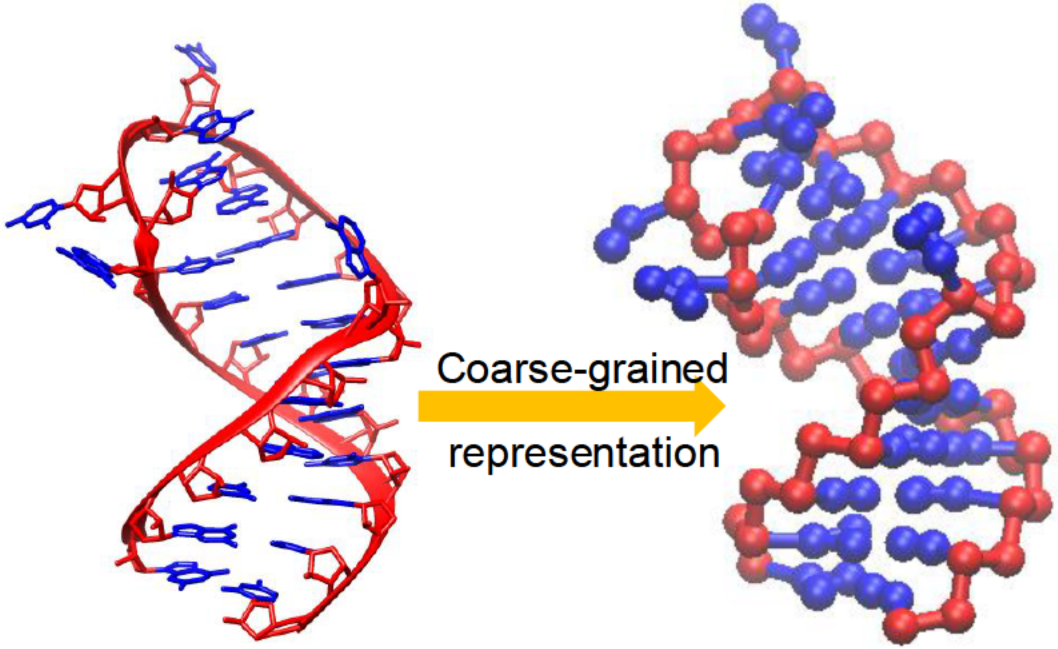


2D structure



3D structure
(Yeast tRNA^{Asp}; PDB code: 1VTQ)

To predict the (final) 3D from (2D+initial 3D)
MD simulation: IsRNA



Local connectivity

$$E_{total} = \boxed{\sum E_{bond}(b) + \sum E_{angle}(\theta) + \sum E_{torsion}(\phi)}$$
$$+ \sum E_{bp}(b, \theta, \phi) + \sum E_{pair}(r) + \sum E_{LJ}(r)$$

Base pairing interactions

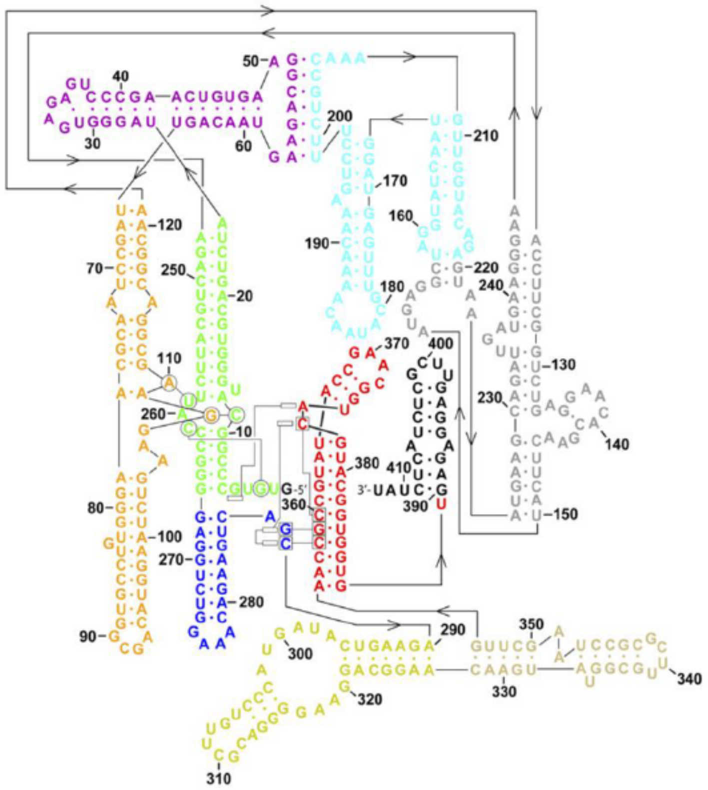
Non-local pairwise interactions

Excluded volume

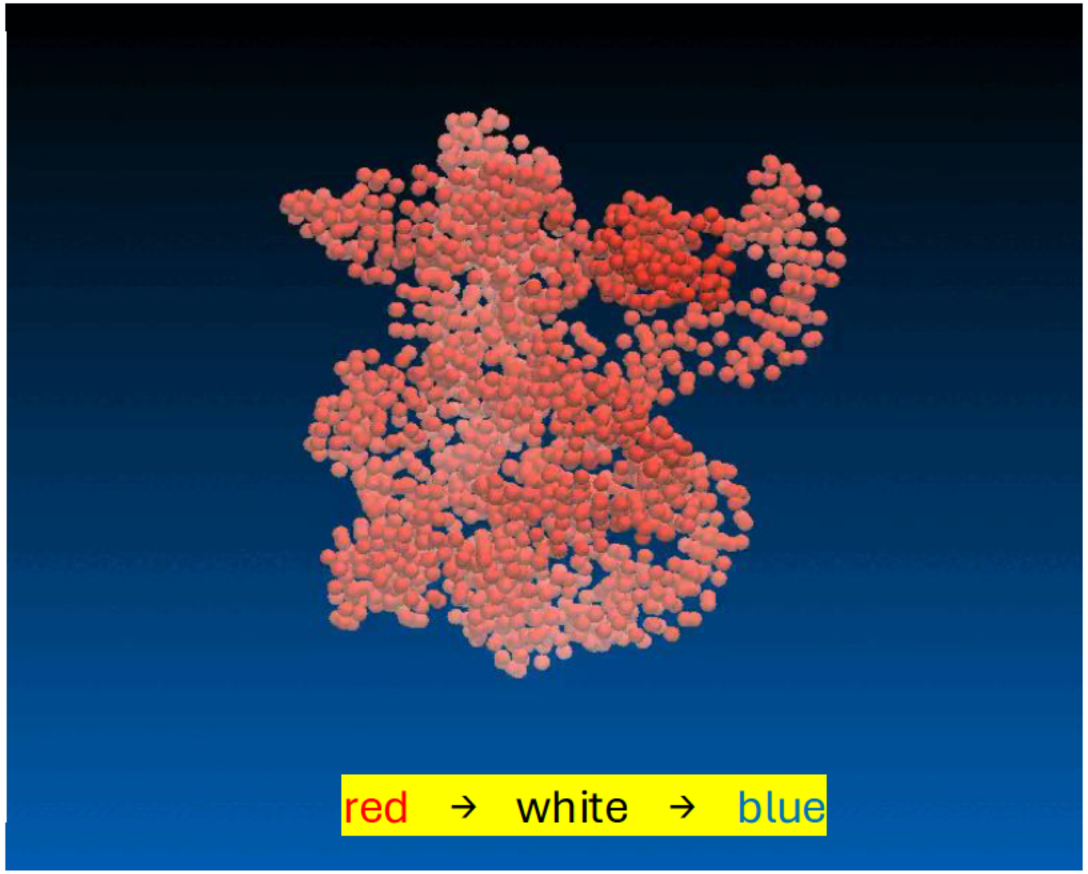
A Bayesian FF derived from known structures

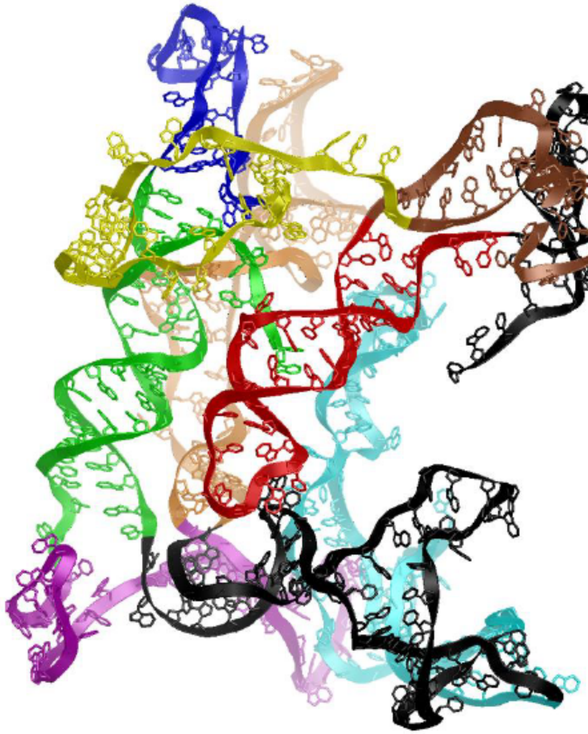
IsRNA: Coarse-grained molecular dynamics simulation

2D structure → **coarse-grained MD simulation** → **2D & 3D structures**



Group II intron (412 nts)

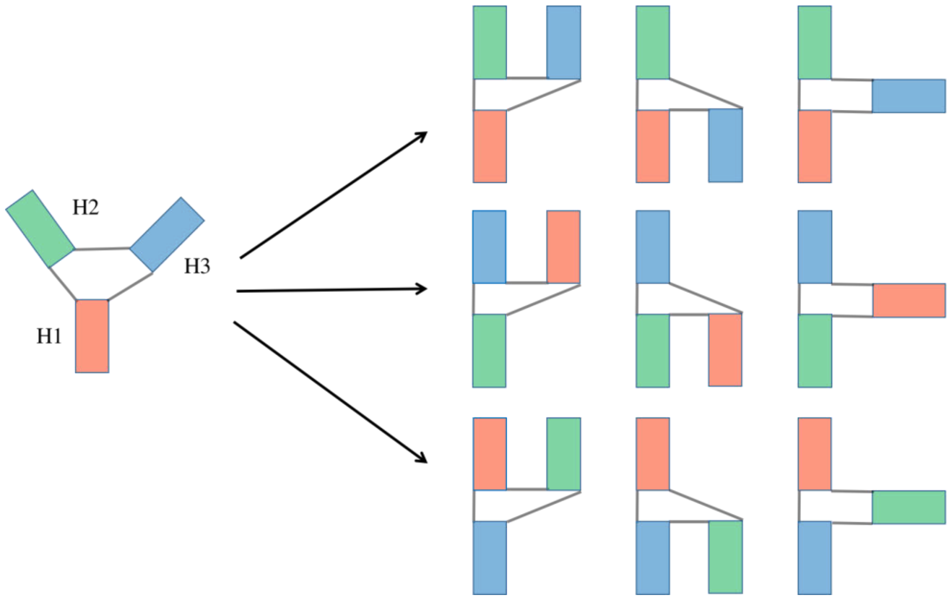


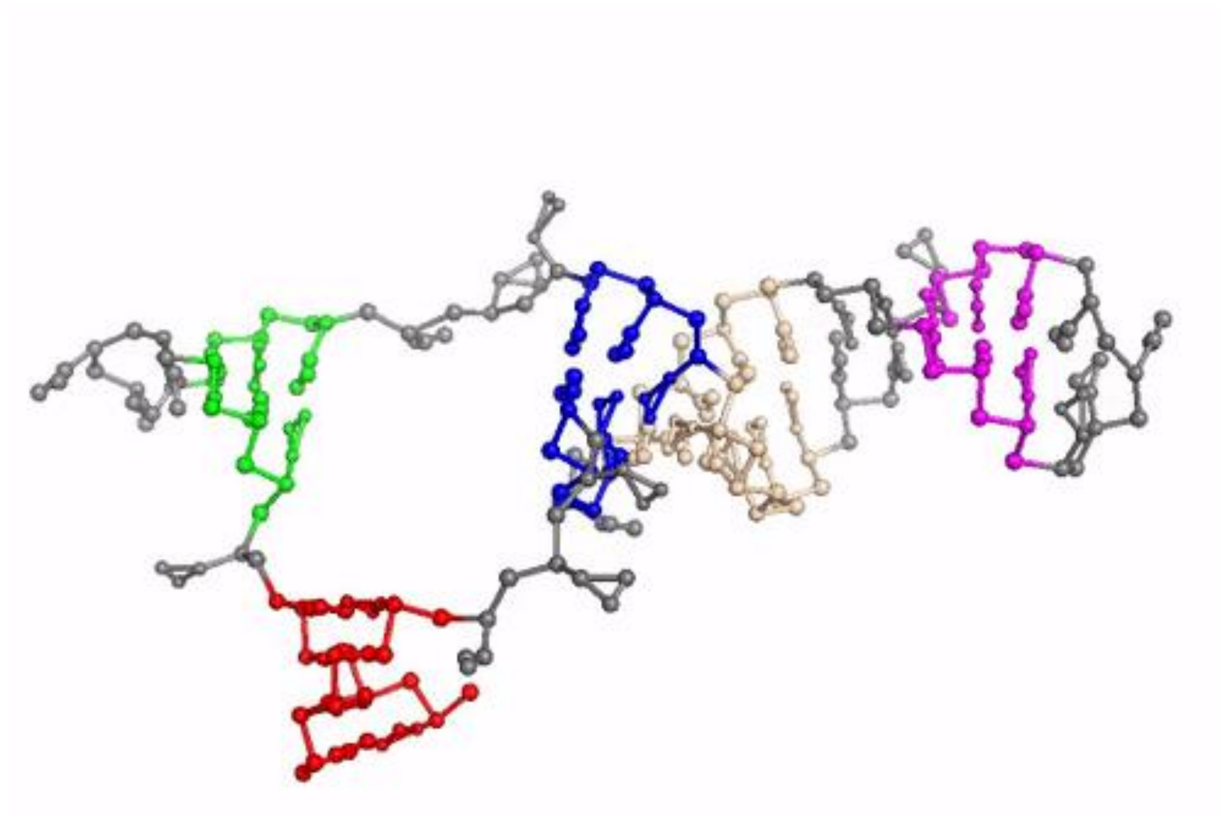


- Can also predict sub-populated alternative folds

To predict the (final) 3D from (2D+initial 3D)

RNAJP: MC/MD hybrid simulation





Li, J., & Chen, S. J., (2022) Nucleic Acids Research

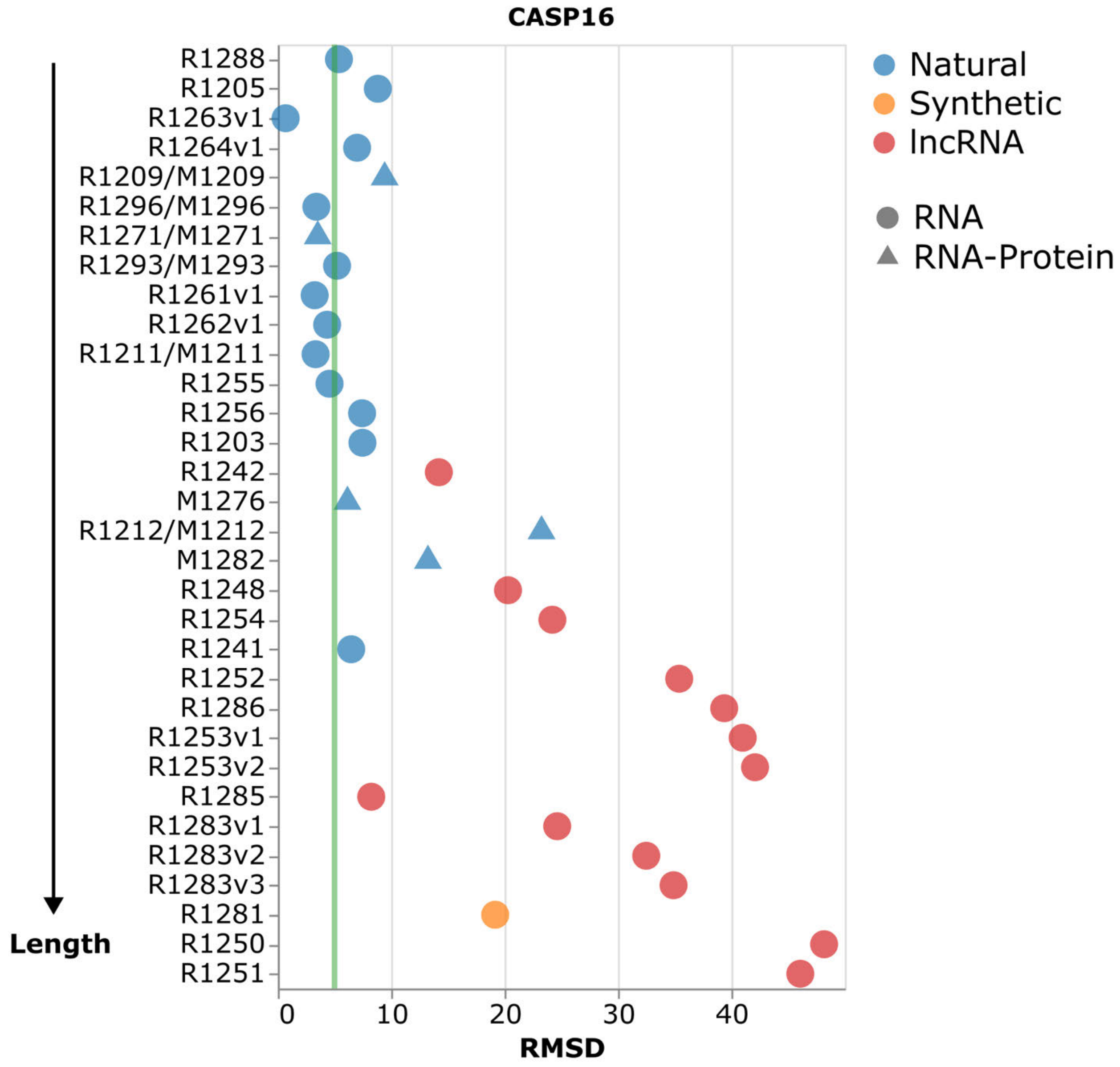
AlphaFold 3

- ❑ If AF3 models agree with the Vfold predictions
  We are happy!

- ❑ If AF3 models are different from the Vfold predictions
  We are also happy: AF3 model + 2D → MD simulation

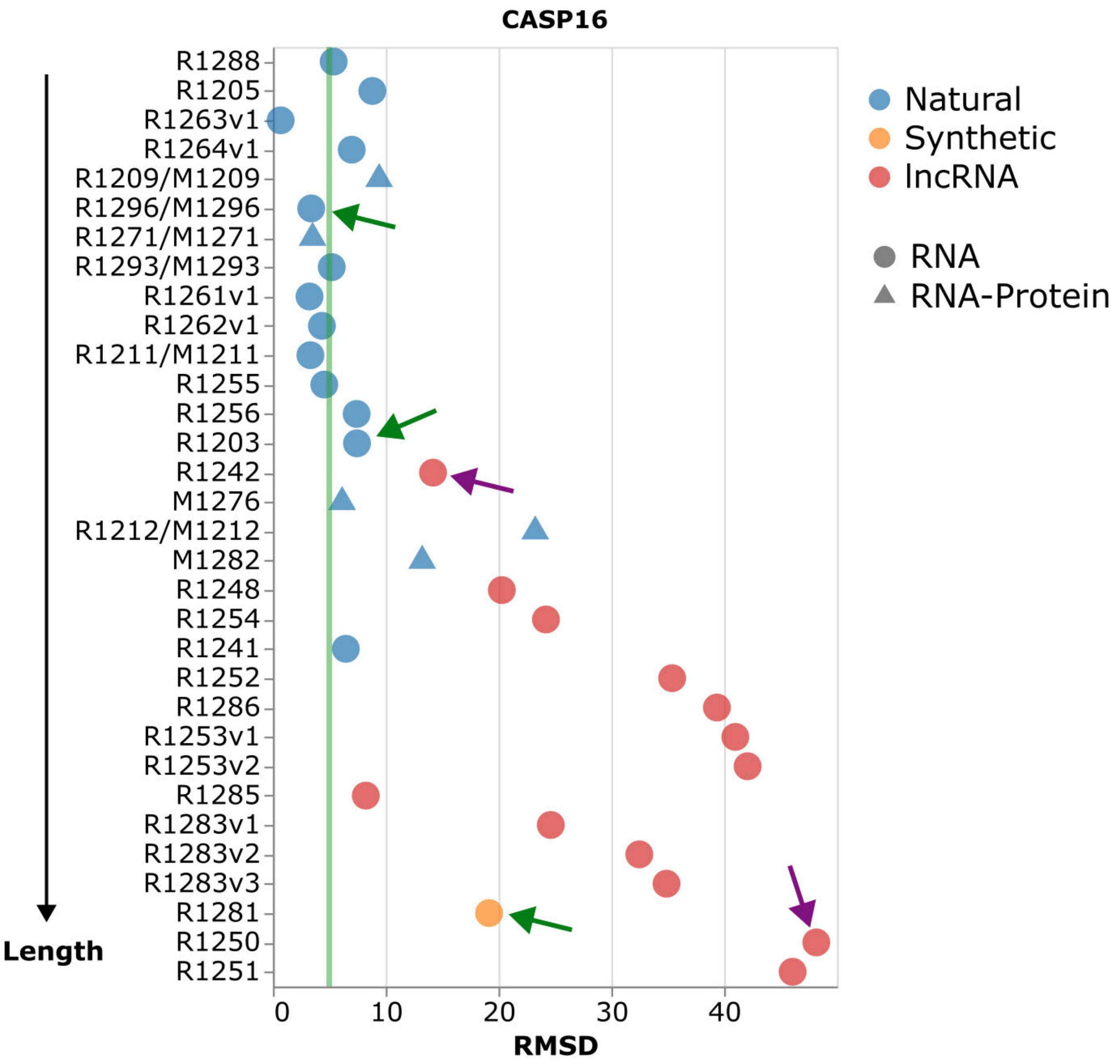
We are always happy with AF3!

Vfold performance in CASP16 – an overview



- **Strong performance** on **small RNAs and natural RNAs** due to the availability of (some) templates and/or the computationally efficient simulations capable of generating extensive structural ensembles for analysis.
- **Poor performance** on **lncRNAs**, as template information and additional structural information is not readily available, resulting in computationally challenging simulations.
- **Mixed performance** on **RNA-protein targets**. Without template or literature information, RNA-protein docking presents a significant challenge.

Five representative targets

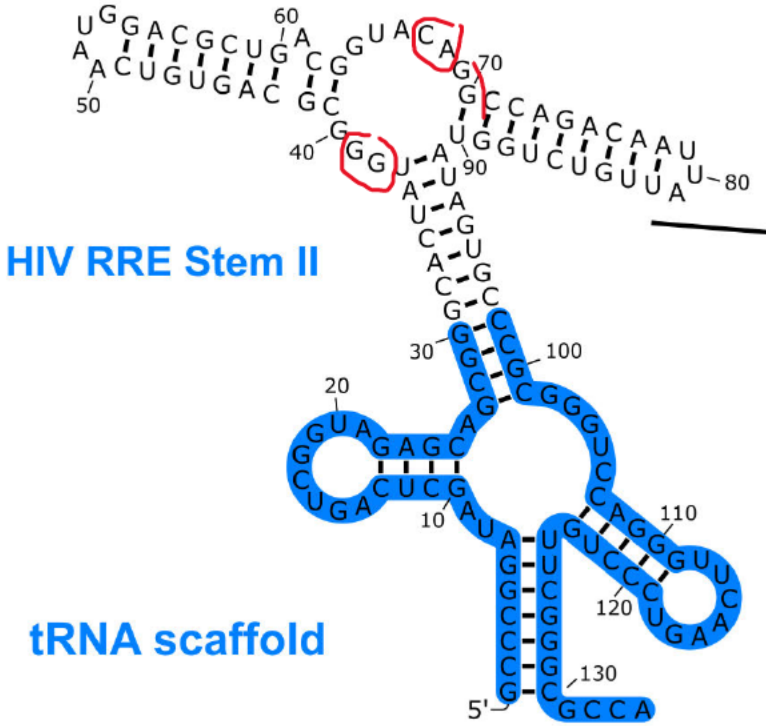


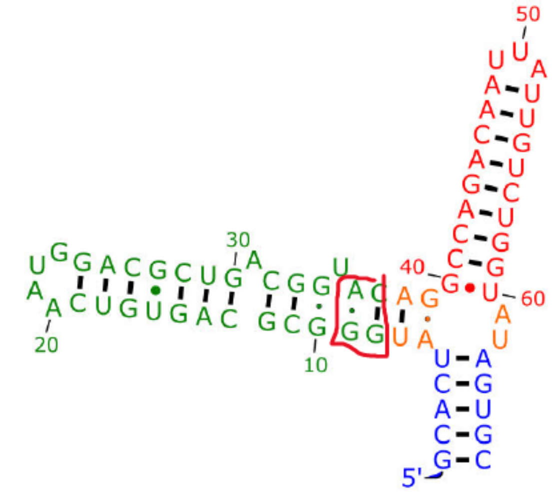
Four representative targets:

The Good:
- Two RRE SL2 targets (our best models: 6.68 Å and 3.40 Å)
- RNA Origami dimer (our best model: 19.17 Å)

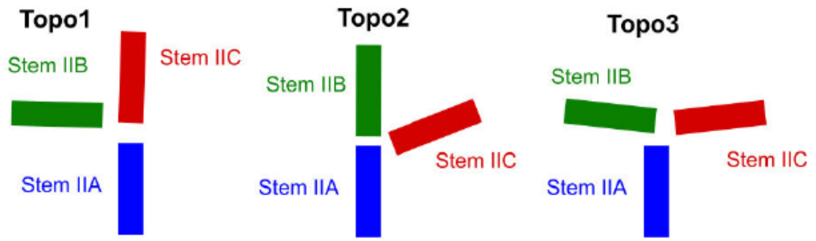
The Bad:
- raiA RNA (our best model: 15.84 Å)
- GOLLD lncRNA (our best model: 48.18 Å)

(1) RRE SLII (R1203): Importance of template refinement

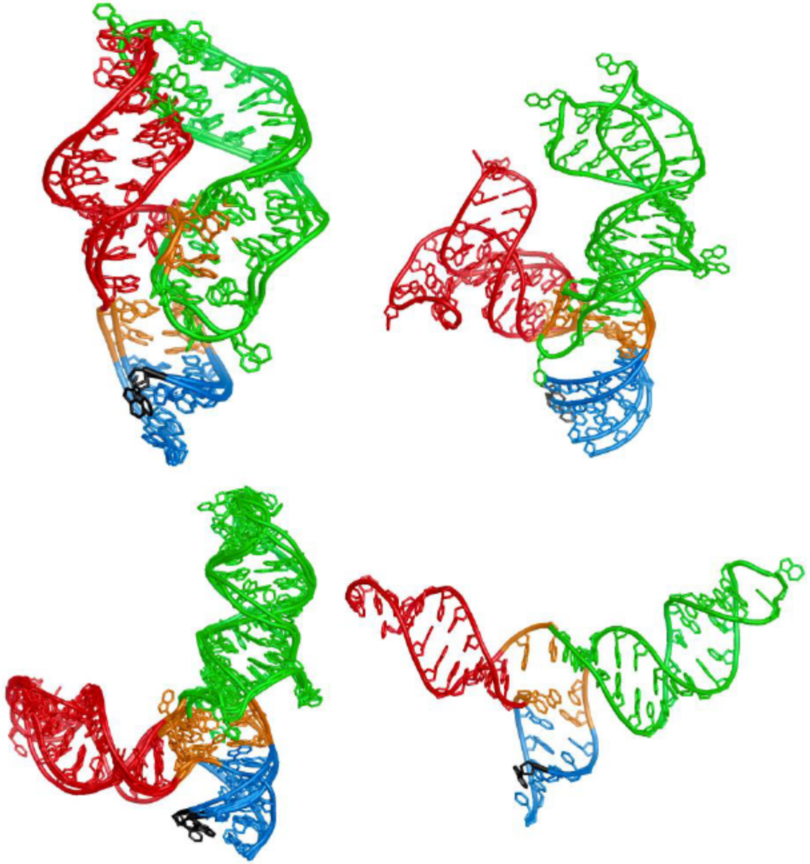




The predicted 2D structure

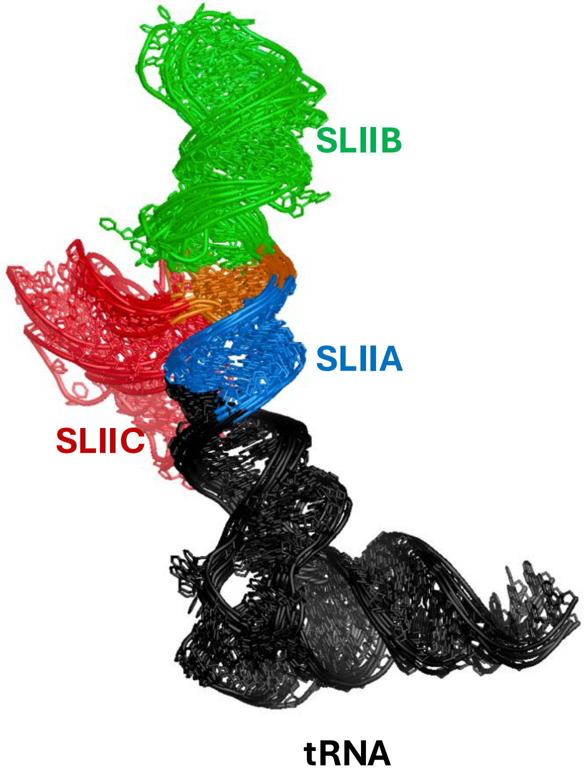


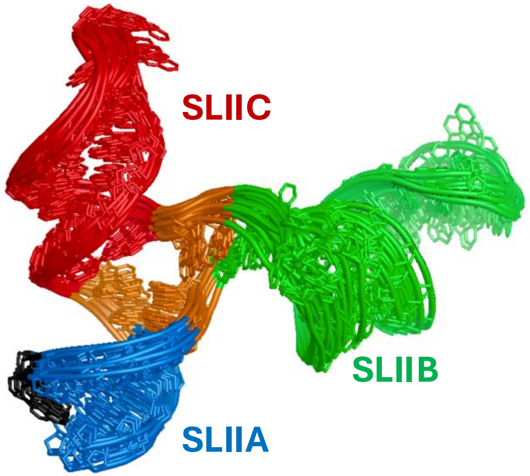
3 potential coaxial stacking modes



Rev response element (**RRE**)+**Rev** protein → transport viral mRNA out of nucleus

AF3 → different T-shaped topologies- w/wt tRNA scaffold

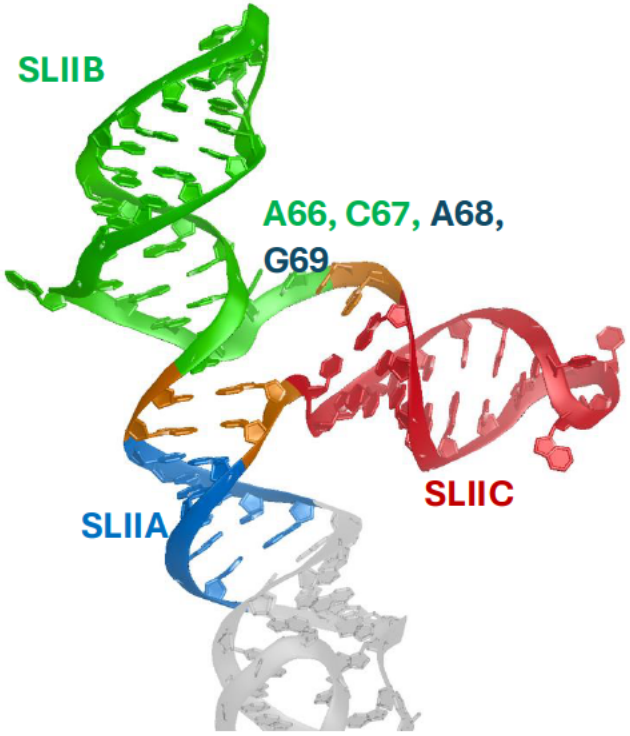




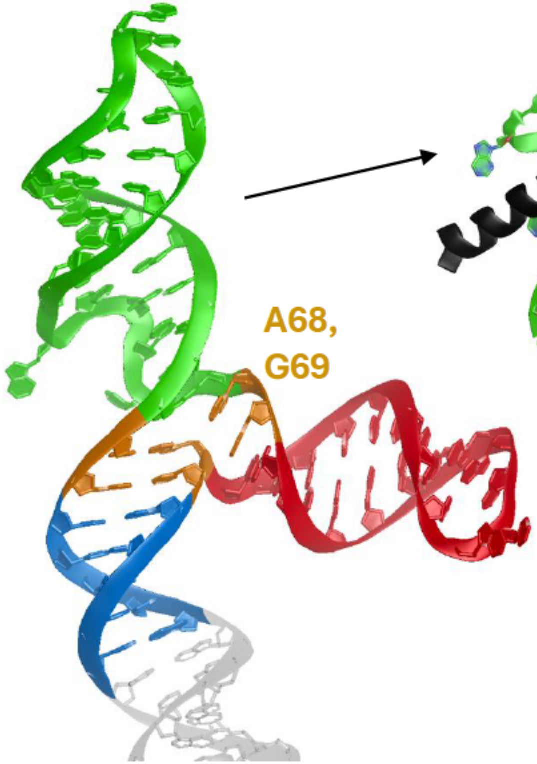
SLII with the tRNA scaffold (R1203): Good

SLII without the tRNA scaffold (R1203): Bad

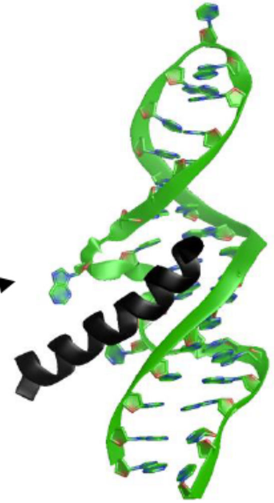
(1) RRE SLII (R1203): Importance of template refinement



The native structure



Our best model: 6.68 Å



The template we used:
RRE SLIIB with Rev peptide (PDB 1ETG)

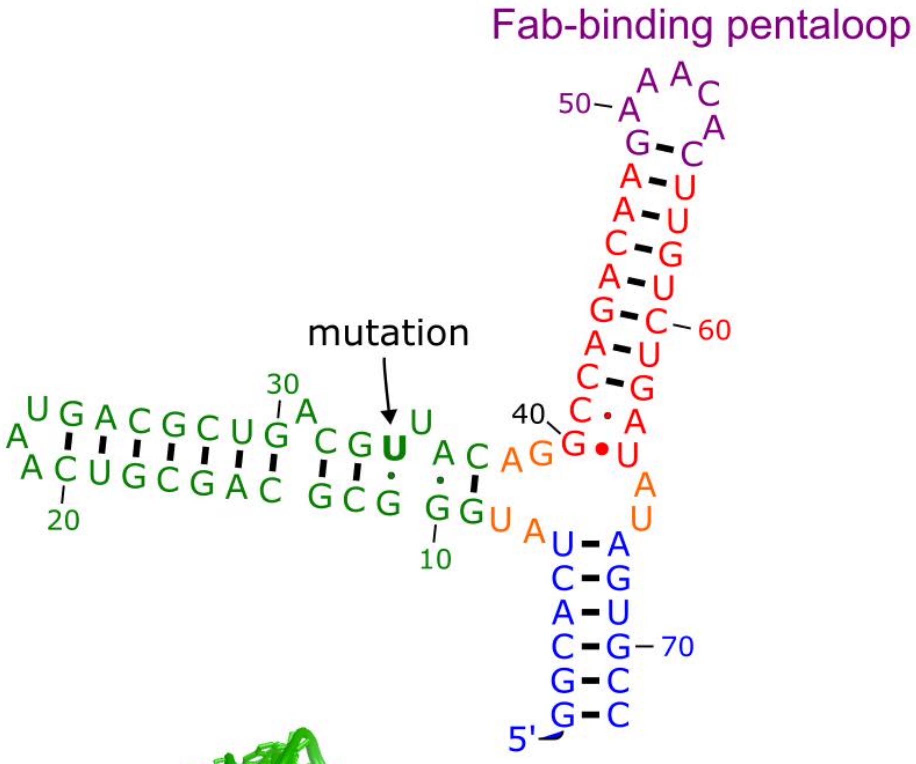
The good: Vfold correctly predicted the coaxial stacking between SLIIA and SLIIB, as well as the T-shape of the 3WJ.

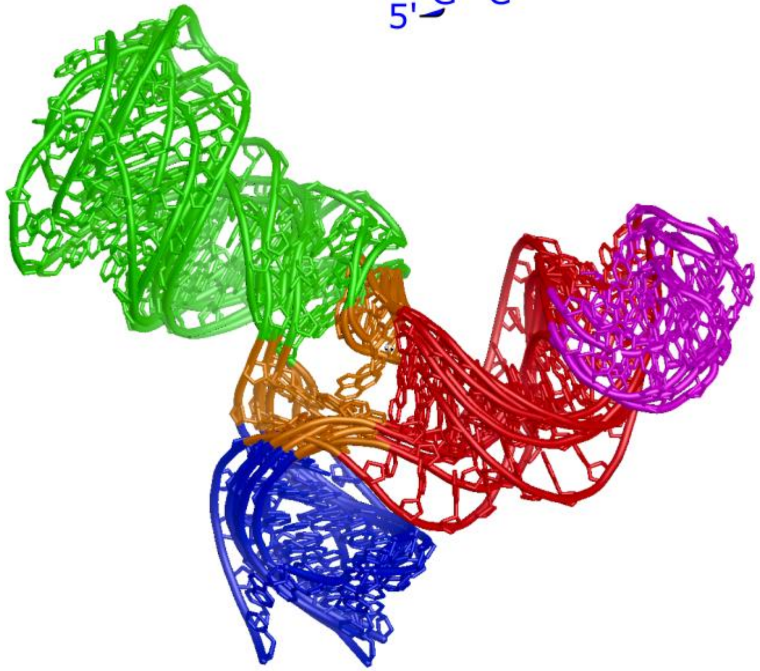
The bad: A66 and C67 form base pairs within SLIIB (following the template), but in the native structure, they flip out and stack with A68 and G69.

(2) RRE SLII (R1296):
AF3-assisted sampling

A single mutation →
Similar 2D but dramatically different 3D

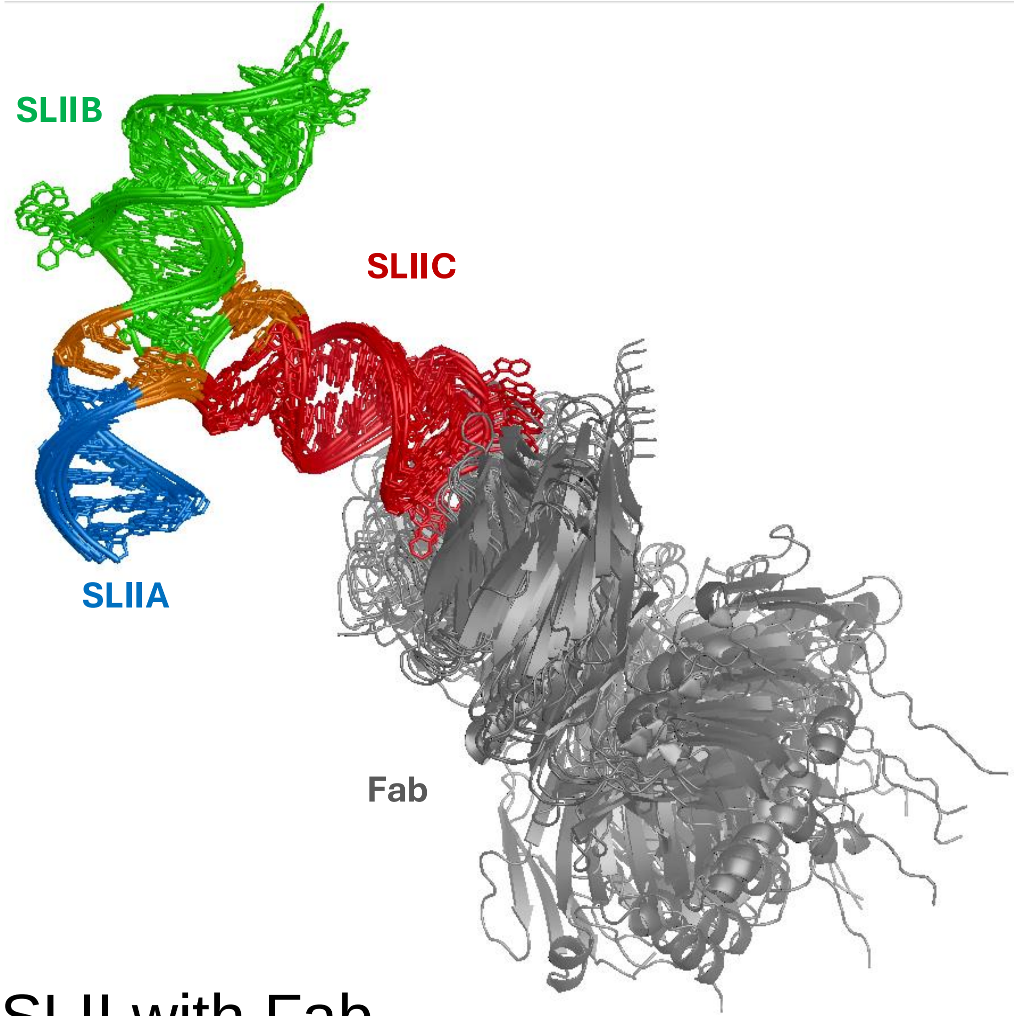
The predicted 2D structure



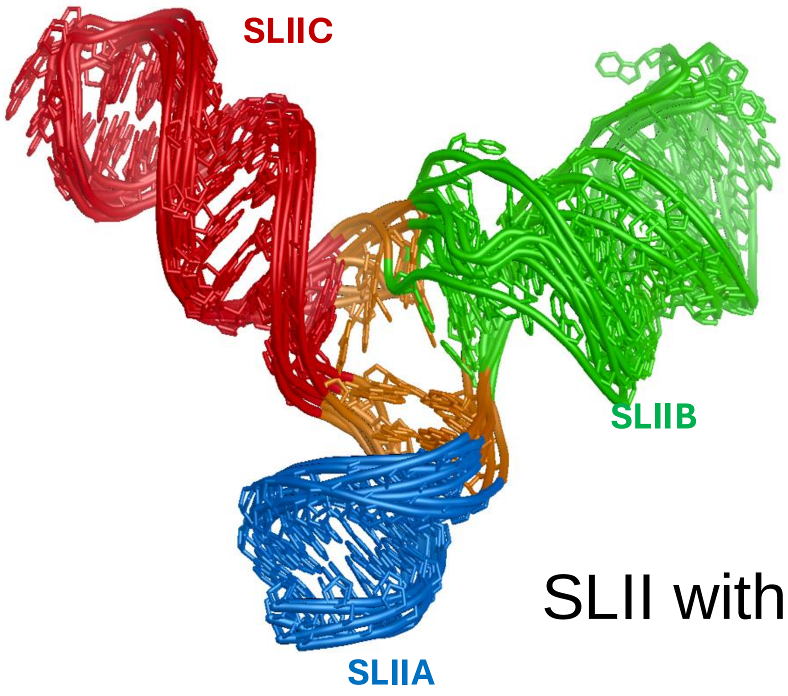


Vfold suggests a highly dynamic 3WJ

AF3 → Y-shaped topologies (Fab-free)



SLII with Fab

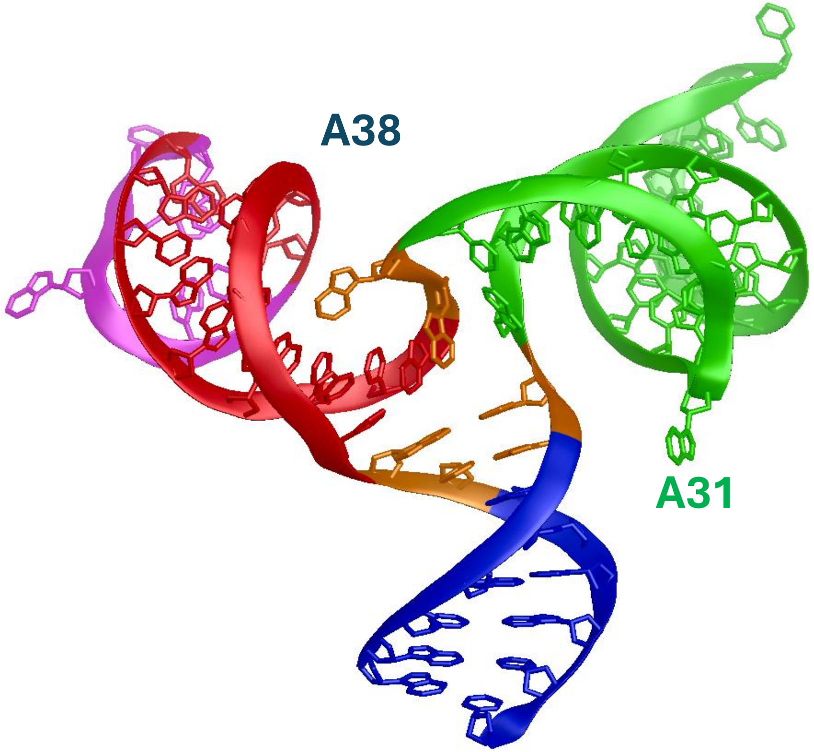


SLII without Fab

Y-shape: No coaxial stacking between SLIIA, SLIIB, SLIIC
Absolutely dominant structure by AF3

(2) RRE SLII (R1296): AF3-assisted sampling

[redacted]



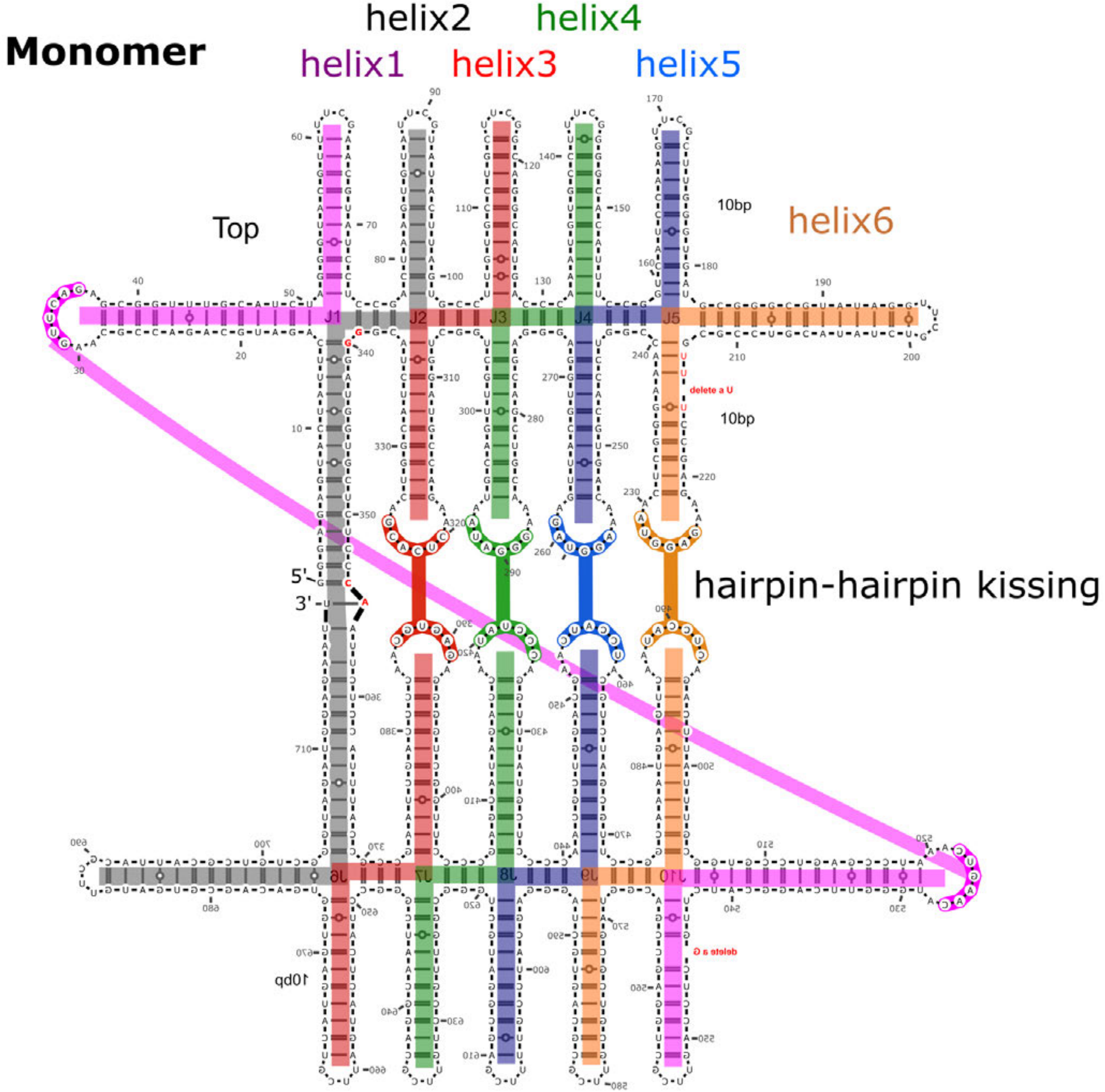
The native structure

Our best model: 3.4 Å

The good: Vfold correctly modeled the 3WJ conformation, in which the three arms do not exhibit coaxial stacking.

The bad: Vfold incorrectly predicts that A38 flip out to form base triples; additionally, we missed the A-minor interaction between A31 and SLIIA.

(3) RNA origami dimer (R1281)

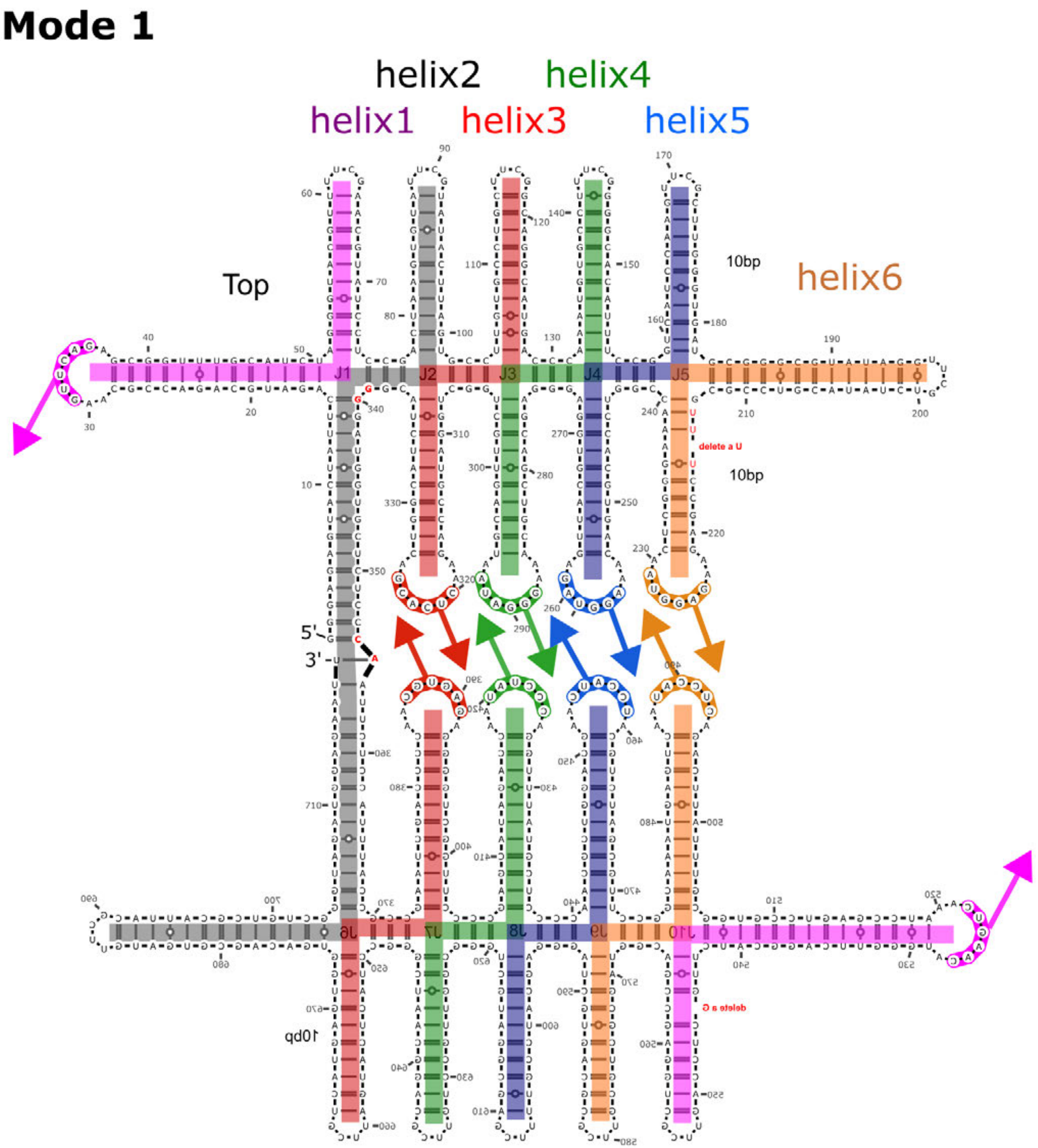


Monomer:
6-helix bundle origami RNA (PDB 7PTK)

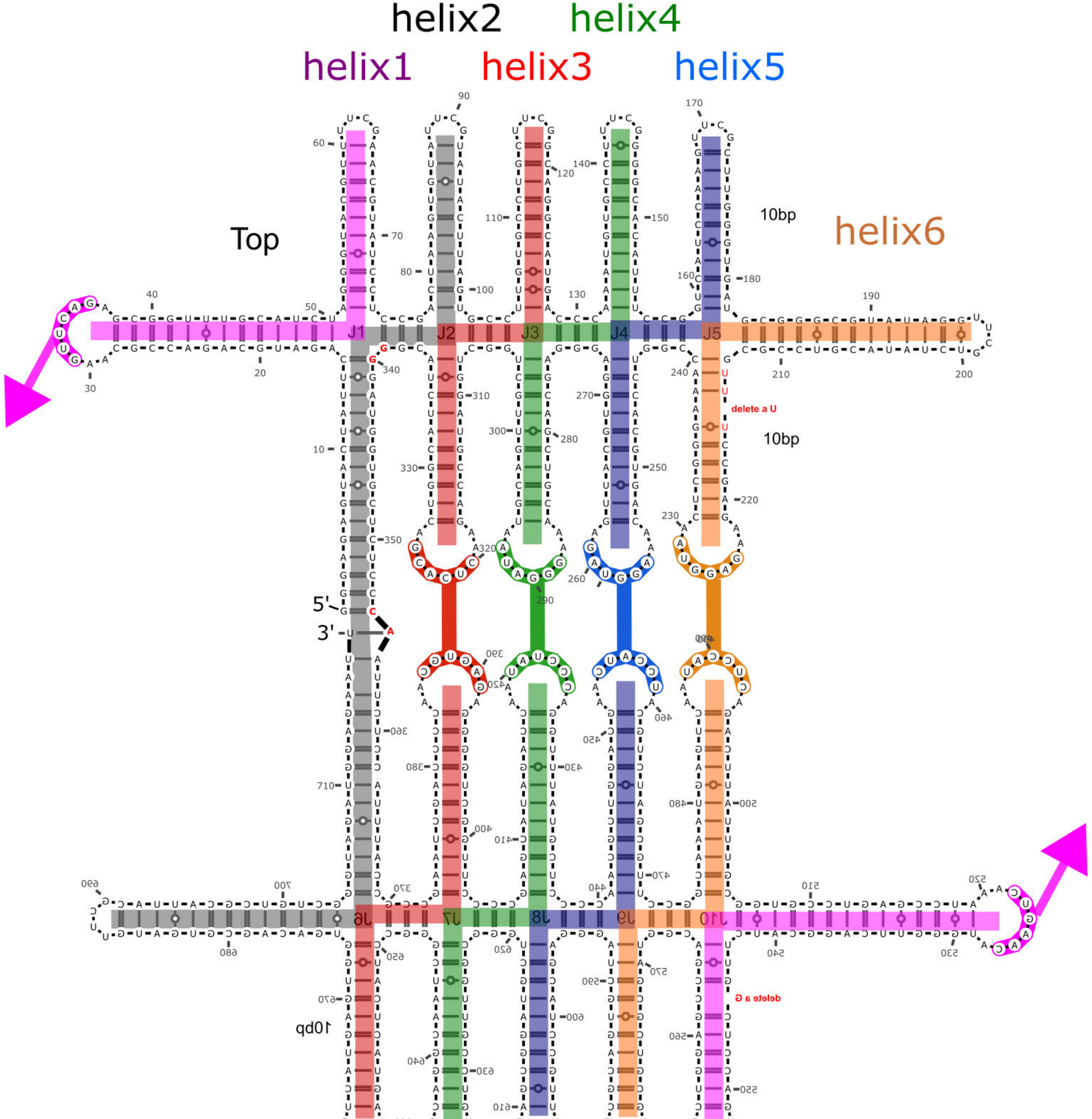
Low-barrier dimerization:

(A1-B1) + (A2-B2) →
(A1-B2) + (A2-B1)

AF3 fails to generate a viable structure

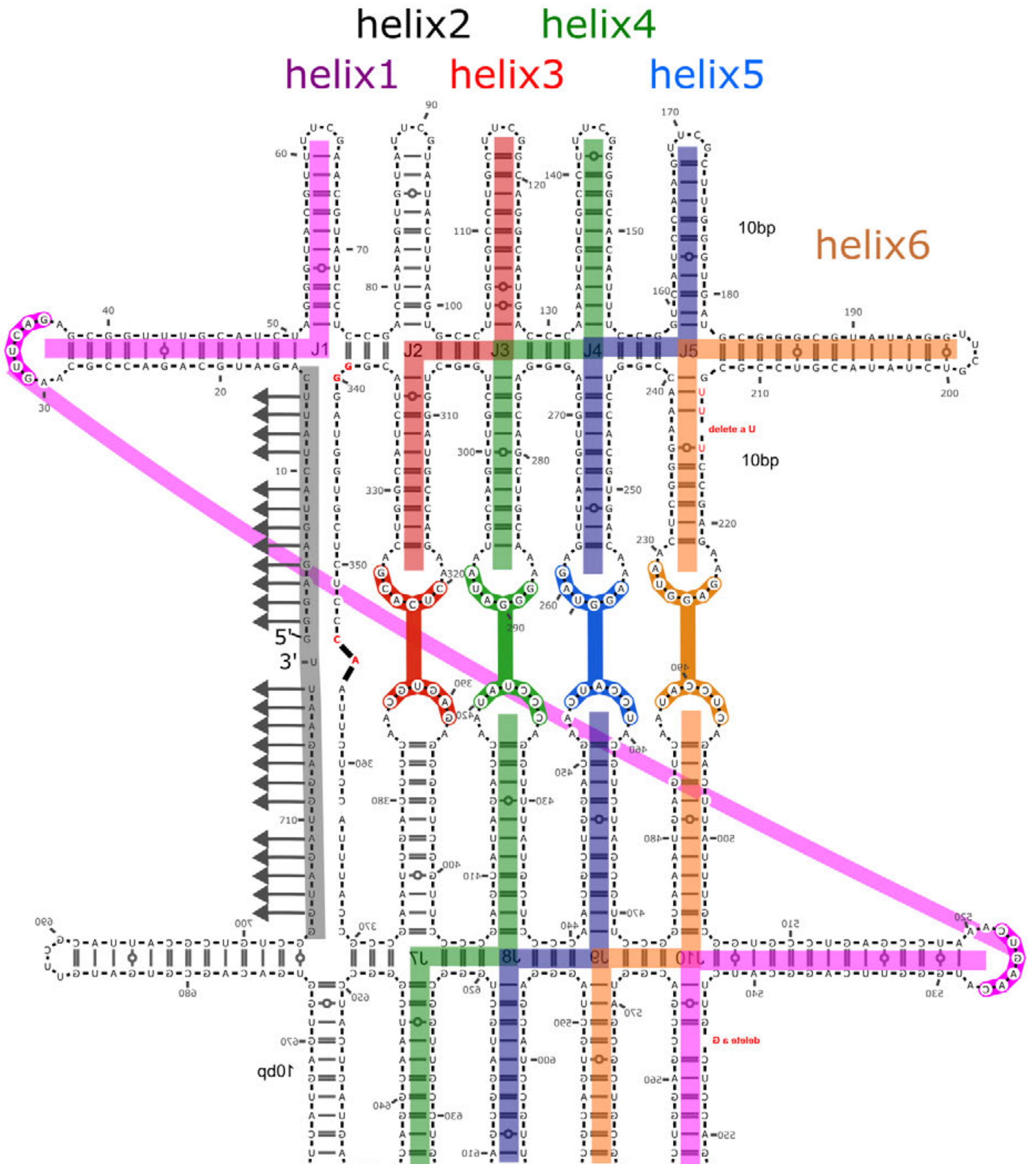


Mode 2



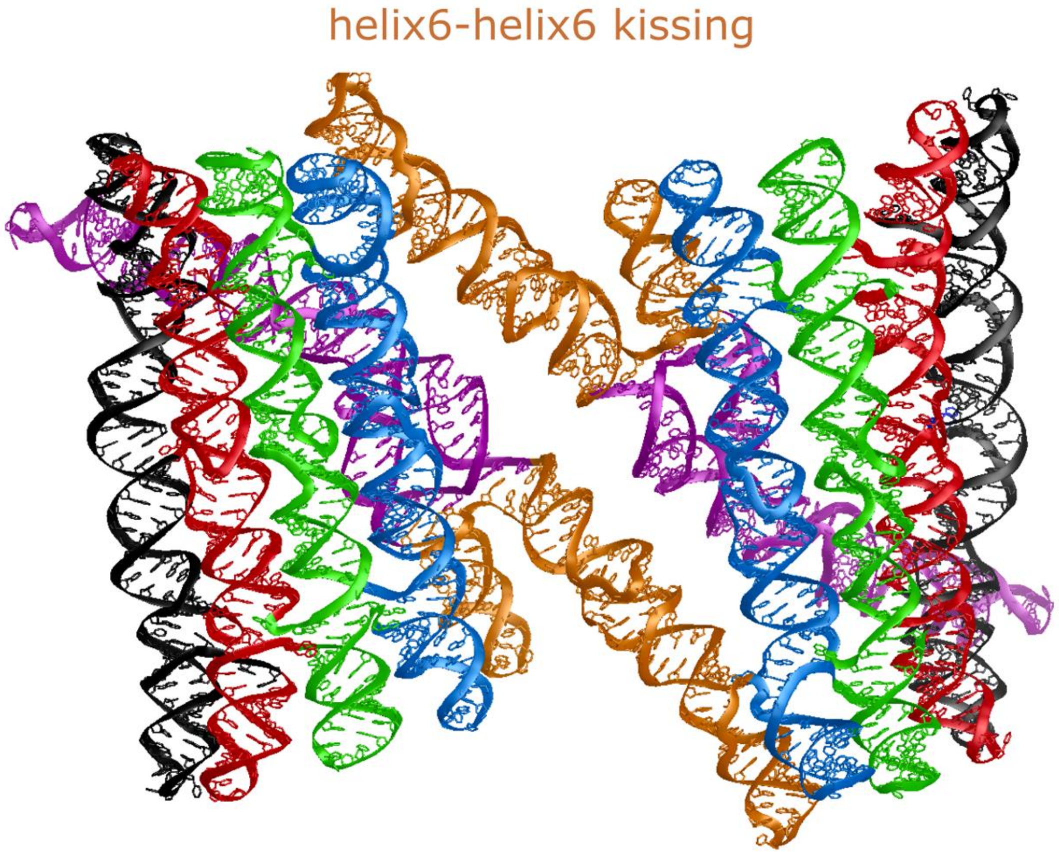
helix2
helix4
helix1
helix3
helix5
Top
10bp
helix6
J1
J2
J3
J4
J5
delete a U
10bp
5'
3'

Mode 3

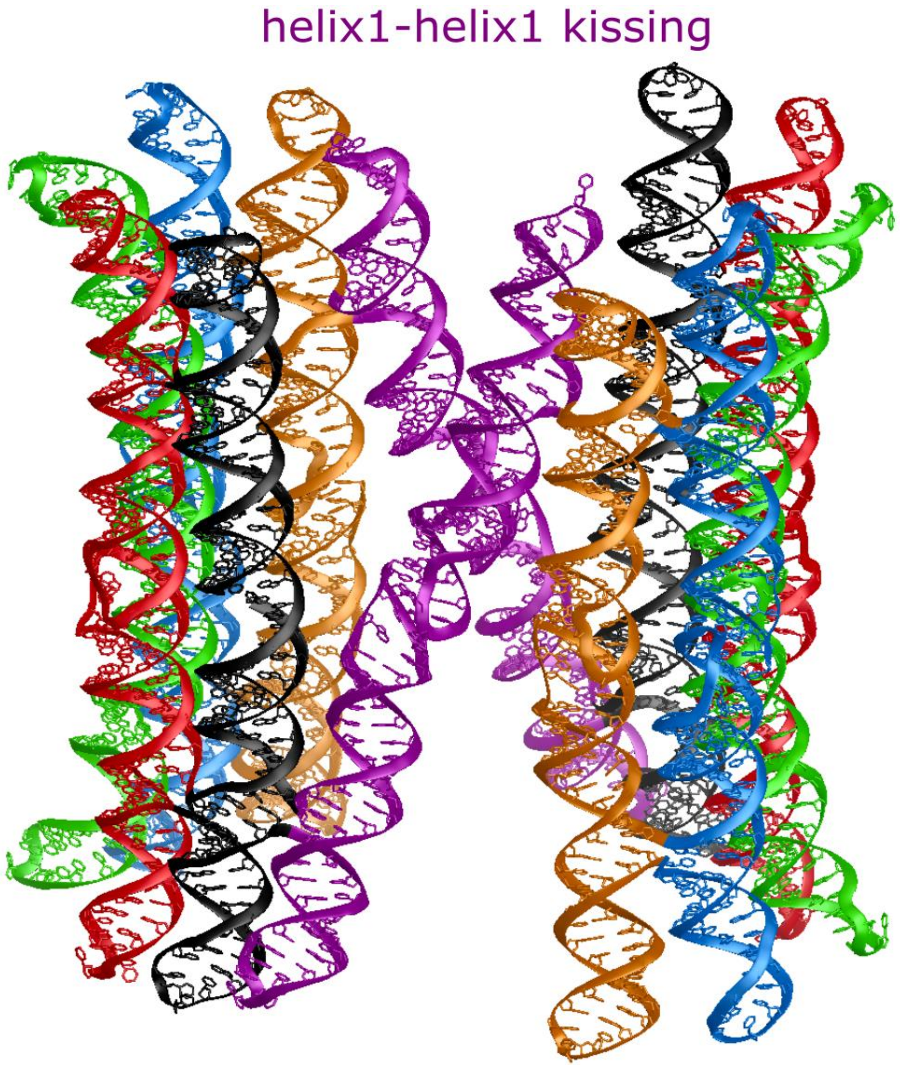


helix2
helix4
helix1
helix3
helix5
10bp
helix6
J1
J2
J3
J4
J5
delete a U
10bp
5'
3'

(3) RNA origami dimer (R1281)



The native structure (dimer)
Mode 2 but for helix 6

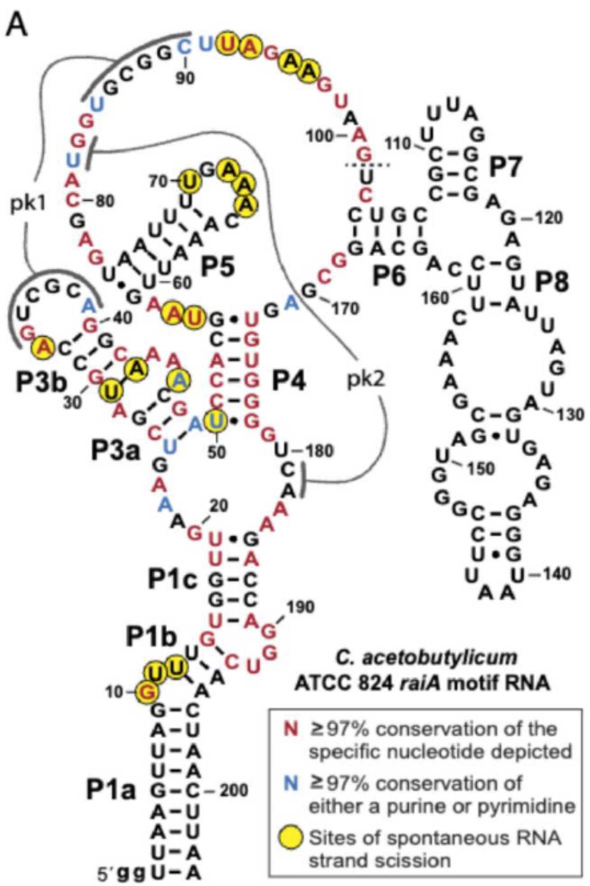


Our best model (dimer) mode 2

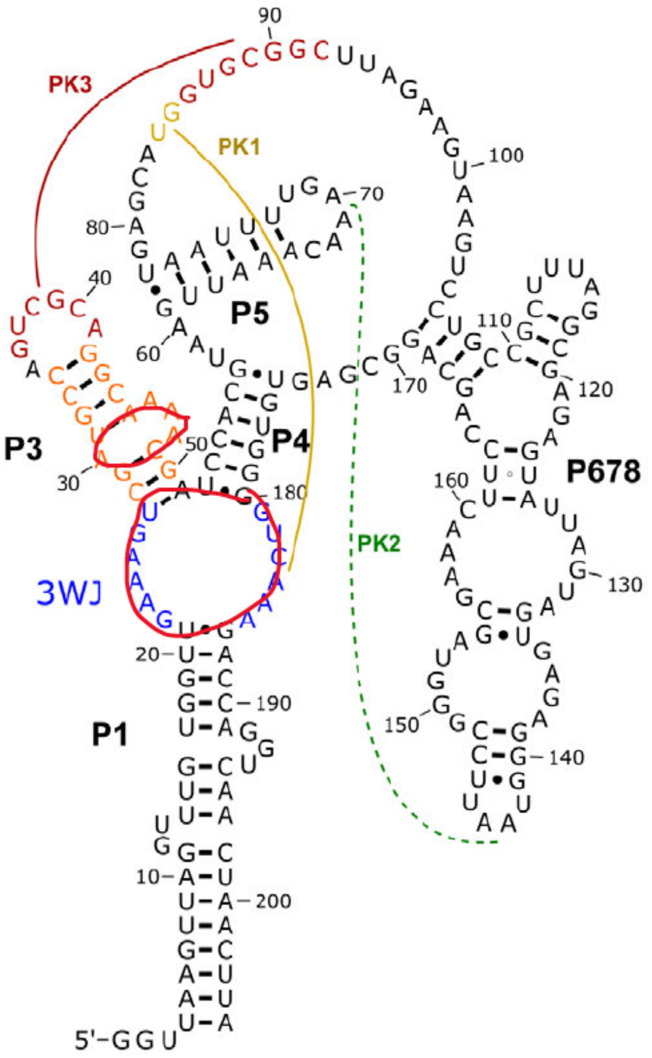
Monomer from our best model (red) against the native structure (green): 19.17 Å

(4) raiA RNA (R1242): good 2D → bad 3D

Structure information and in-line probing experiment on raiA motif from literature[1].



The 2D structure was predicted by combining literature info with our models.



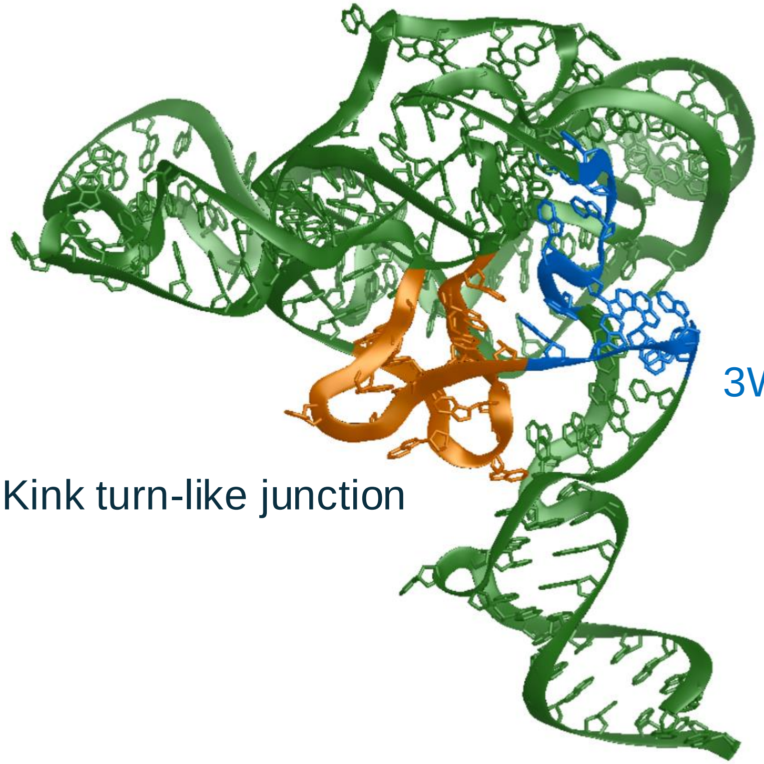
The good: Vfold predicts good 2D structure

Specificity: 0.88 (ppv)
Sensitivity: 0.94 (recall)
F1-score: 0.91

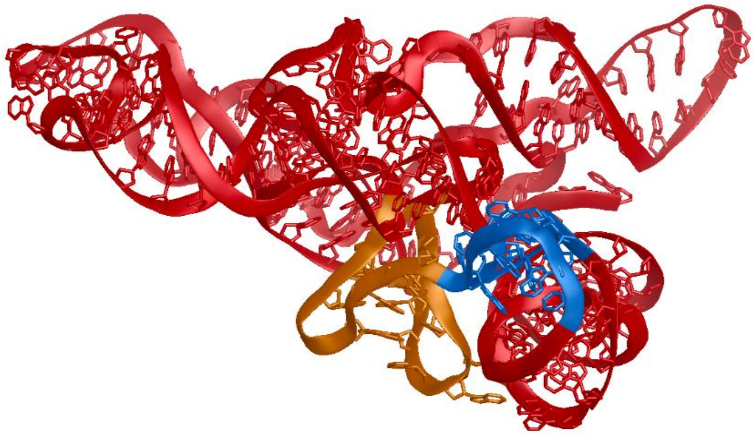
The three pseudoknots are all correctly predicted.

[1] Soares, Lucas W., Christopher G. King, Chrishan M. Fernando, Adam Roth, and Ronald R. Breaker. (2024). *PNAS*

(4) raiA RNA (R1242): Good 2D → bad 3D



The native structure



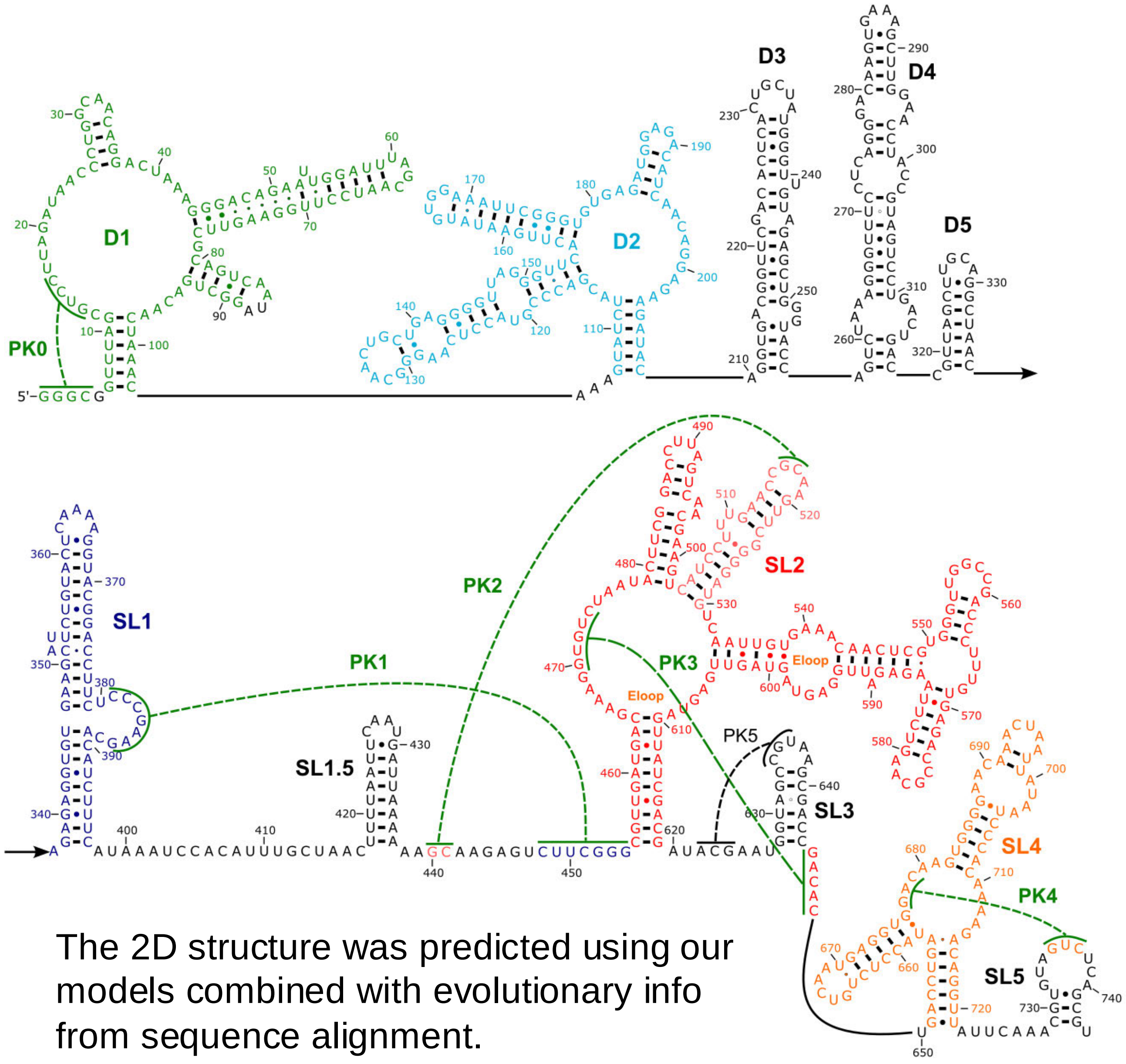
Our model: 15.84 Å

The bad: the kink turn-like junction and the large 3WJ

Vfold fails to generate the initial 3D structure

AF3 generated initial 3D fold + 2D structure → the predicted model

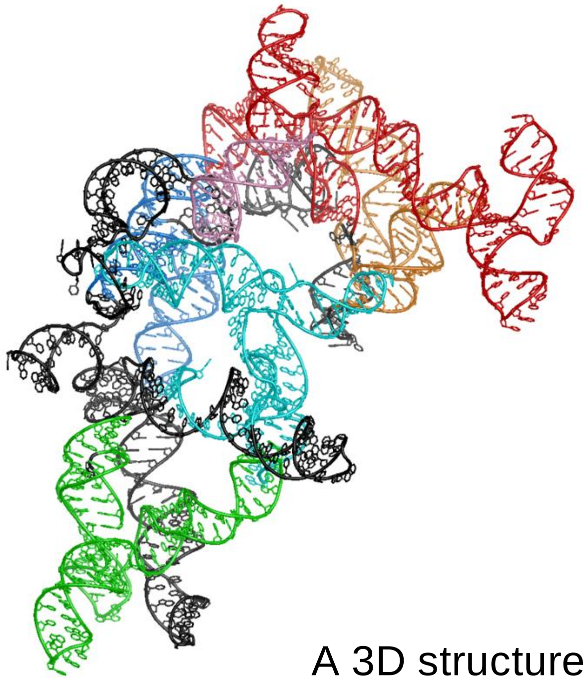
(5) GOLLD lncRNA (R1250): Structural rearrangement upon R-R binding



The 2D structure was predicted using our models combined with evolutionary info from sequence alignment.

The bad: **2D structure prediction accuracy**

Specificity: 0.55 (many false positives)
Sensitivity: 0.85
F1-score: 0.67
The formation of pseudoknots are correctly predicted, but with low specificity.

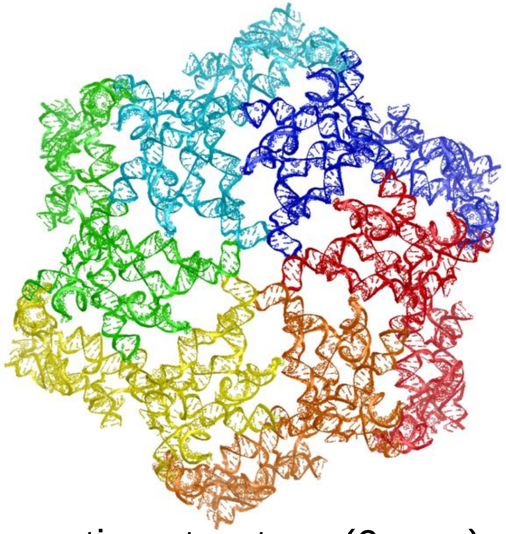


A 3D structure we assembled for simulation

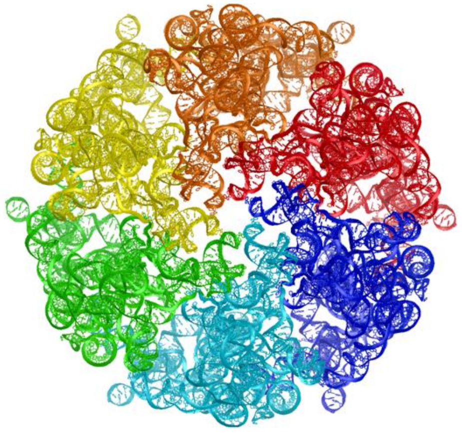
(5) GOLLD lncRNA (R1250): Structural rearrangement upon R-R binding

The native structure

Our best model: 48.18 Å



The native structure (6-mer)



Our best model (6-mer): 67.2 Å

The bad: We first modeled the monomer structure and subsequently assembled the 6-mer structure through RNA-RNA docking. This resulted in overly compact monomer and 6-mer structures.

This target highlights the significant challenges in predicting large and multimeric RNA structures at both the 2D and 3D levels.

Perspectives

What we did right:

- Natural and small RNAs
- Conformation sampling using CGMD simulations
- Prediction of dimerization modes

What went wrong:

- Large RNA, both at 2D and 3D levels
- Multimeric RNAs: docking of monomeric RNA doesn’t work
- RNA-Protein complexes

Future directions:

- New models for the simulations of large/multimeric RNAs
- Integration of machine learning-base models with simulation methods
- Improving RNA-protein docking, accounting for RNA flexibility

Acknowledgment

- **Sicheng Zhang**
- Jun Li (GBU)
- Yuanzhe Zhou

- Xiaoqin Zou
- Rui Duan
- Limin Qiu

NIH R35-GM134919
NSF CHE-2154924
NIH U54-AI170660
NIH R01-NS131416