

TEMPLATE-BASED PREDICTION OF RNA TERTIARY STRUCTURE “TROOPER”

Rastislav Galvánek

A decorative graphic consisting of several parallel white lines of varying lengths, slanted diagonally from the bottom-left towards the top-right, located on the right side of the slide.

- ▶ Experimental evaluation of 3D structure of RNA is hard and expensive (crystallography)
- ▶ The number of known RNA sequences is growing
- ▶ 3D RNA structure is related to its function
- ▶ 3D structure is determined by sequence

WHY PREDICTION?



- ▶ De novo methods
 - ▶ Directly from the sequence
 - ▶ Only small structures
- ▶ Comparative modeling methods
 - ▶ From structure of template molecule
 - ▶ Commonly used for proteins

HOW TO PREDICT?

A decorative graphic consisting of several parallel white lines of varying lengths, slanted upwards from left to right, located in the bottom right corner of the slide.

- ▶ Comparative modeling approach
- ▶ Template with similarity at least 60% and known tertiary structure
- ▶ No manual processing
- ▶ Python (Biopython), Shell, Metacentrum (distributed computing infrastructure for academic centers)

TROOPER

A decorative graphic consisting of several parallel white lines of varying lengths and orientations, located in the bottom right corner of the slide.

- Identification of conserved subsequences

- Sliding window

- Process gaps

- Mapping aligned target structure to template sequence

- EMBOSS Needle

1. ALIGNMENT

```

=====
#
# Aligned_sequences: 2
# 1: TARGET
# 2: TEMPLATE
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 175
# Identity:   148/175 (84.6%)
# Similarity: 148/175 (84.6%)
# Gaps:       15/175 ( 8.6%)
# Score: 667.5
#
#
=====
TARGET      1  GGCCGACGGAGGCGCGCCCCGAGAUGAGUAGGCUGUCCCAUCAGGGGAGGA      50
            .| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TEMPLATE    1  ---GGACGGAGGCGCGCCCCGAGAUGAGUAGGCUGUCCCAUCAGGGGAGGA      47

TARGET      51  AUCGGGGACGGCUGAAAGGCGAGGGCGCCGAAGGGUGCAGAGUCCUCCC      100
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TEMPLATE    48  AUCGGGGACGGCUGAAAGGCGAGGGCGCCGAAGCGAGCAGAGUCCUCCC      97

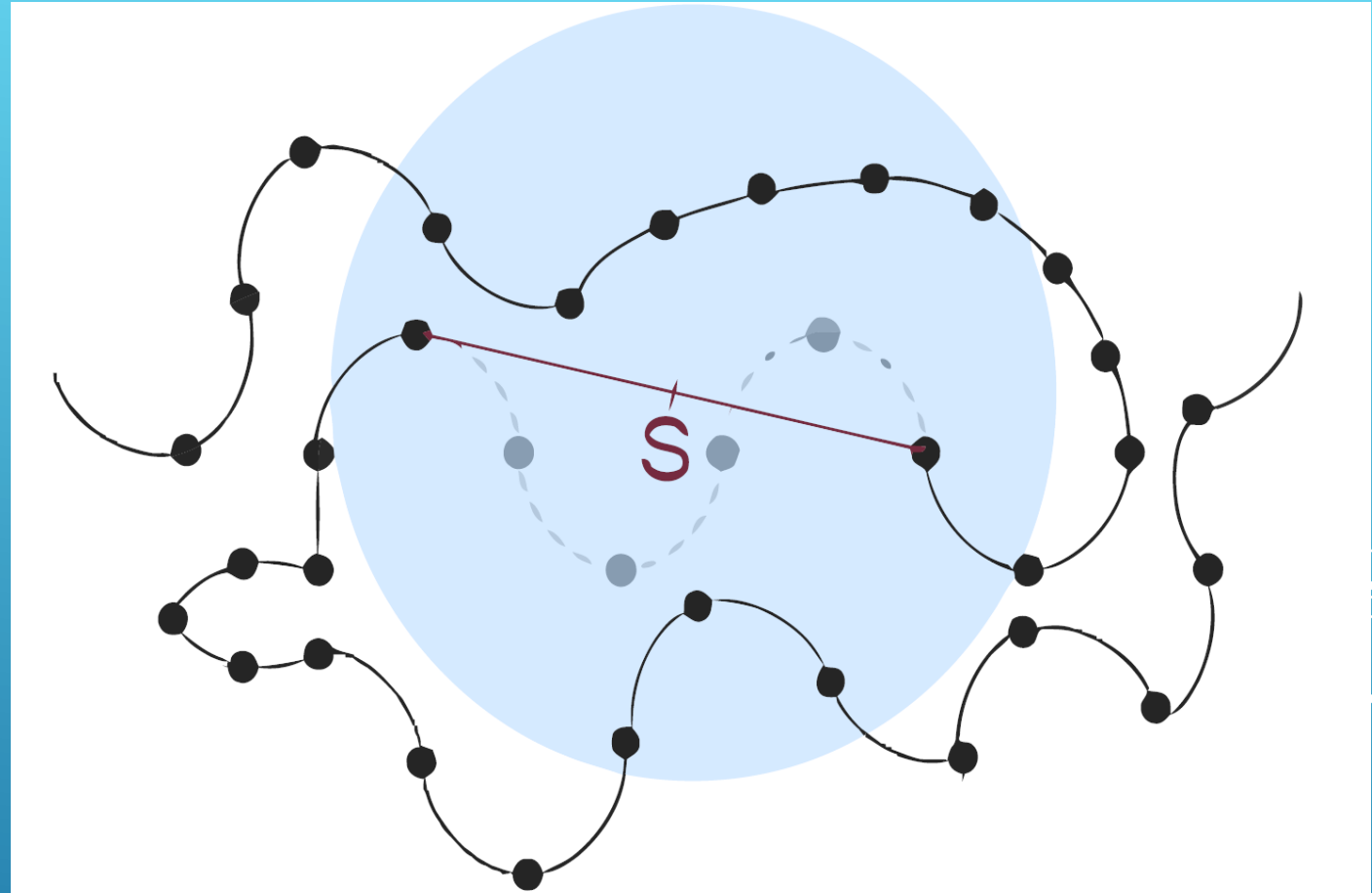
TARGET      101 GCUCUGCAUGCCUGGGGGUAUGGGGAAUACCCAUACCACUGUCACGGAGG      150
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TEMPLATE    98  GCUCUGCUUGGCUGGGGGUGAGGGGAAUACCCUACCACUGUCGCAGAA--      145

TARGET      151 UCUCUCCGUGGAGAGCCGUCGGUC-      174
            . | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TEMPLATE    146 -----AGCGGAGAGCCGUC---CA      161
  
```

- ▶ Comparative modeling
- ▶ RNA Fold from Viena RNA package
- ▶ Used for de novo prediction of 3D structure

2. SECONDARY STRUCTURE REDICTION

- ▶ Special treatment of long unconserved subsequences
- ▶ Avoid collisions
- ▶ Give information



3. PREPARATION FOR DE NOVO PREDICTION

- ▶ FARFAR tool from Rosetta framework
- ▶ Time consuming
- ▶ Poor performance with long sequences

4. DE NOVO PREDICTION

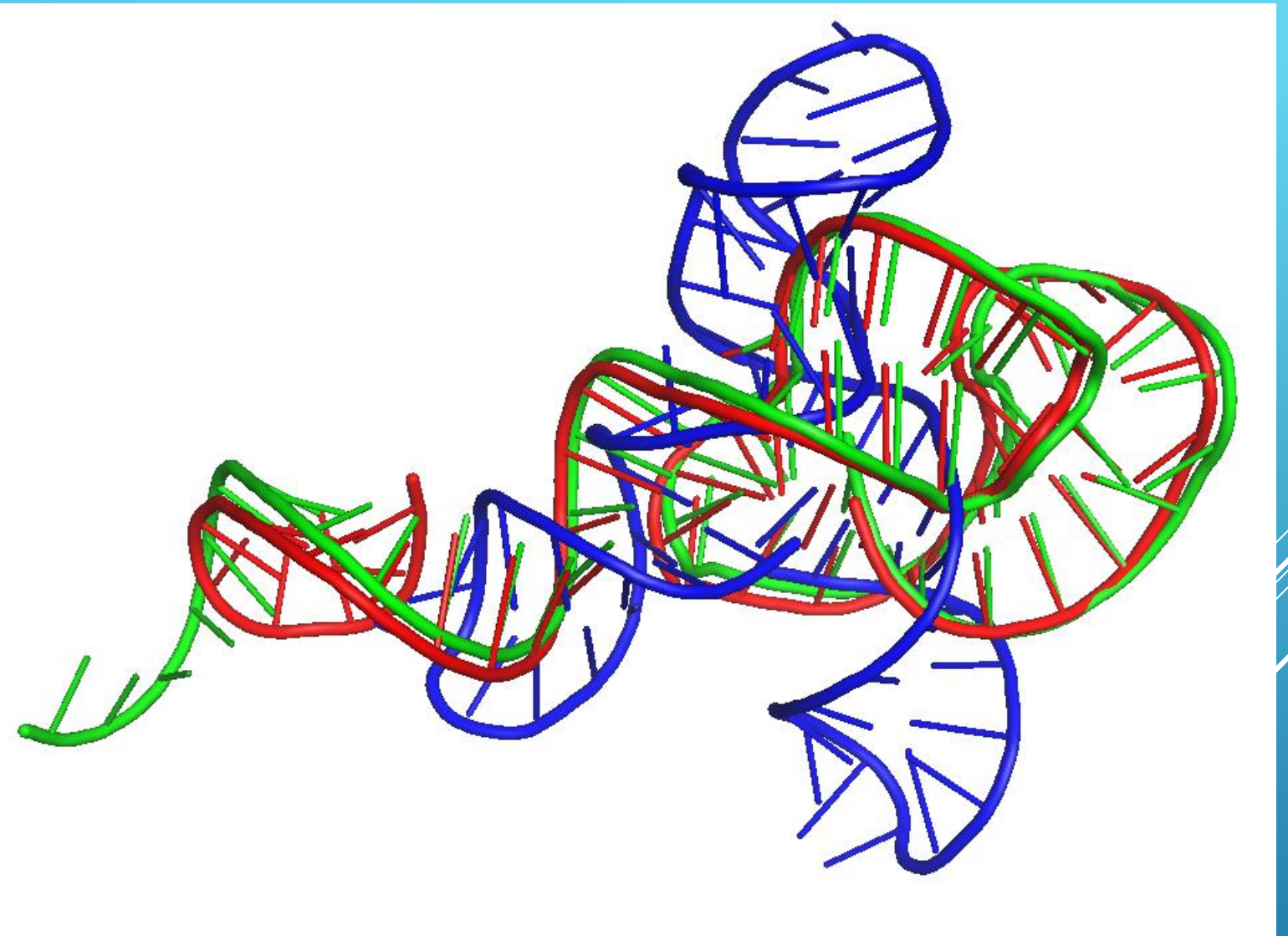


- ▶ Remove duplicities

5. JOINING PREDICTED SUBSTRUCTURES

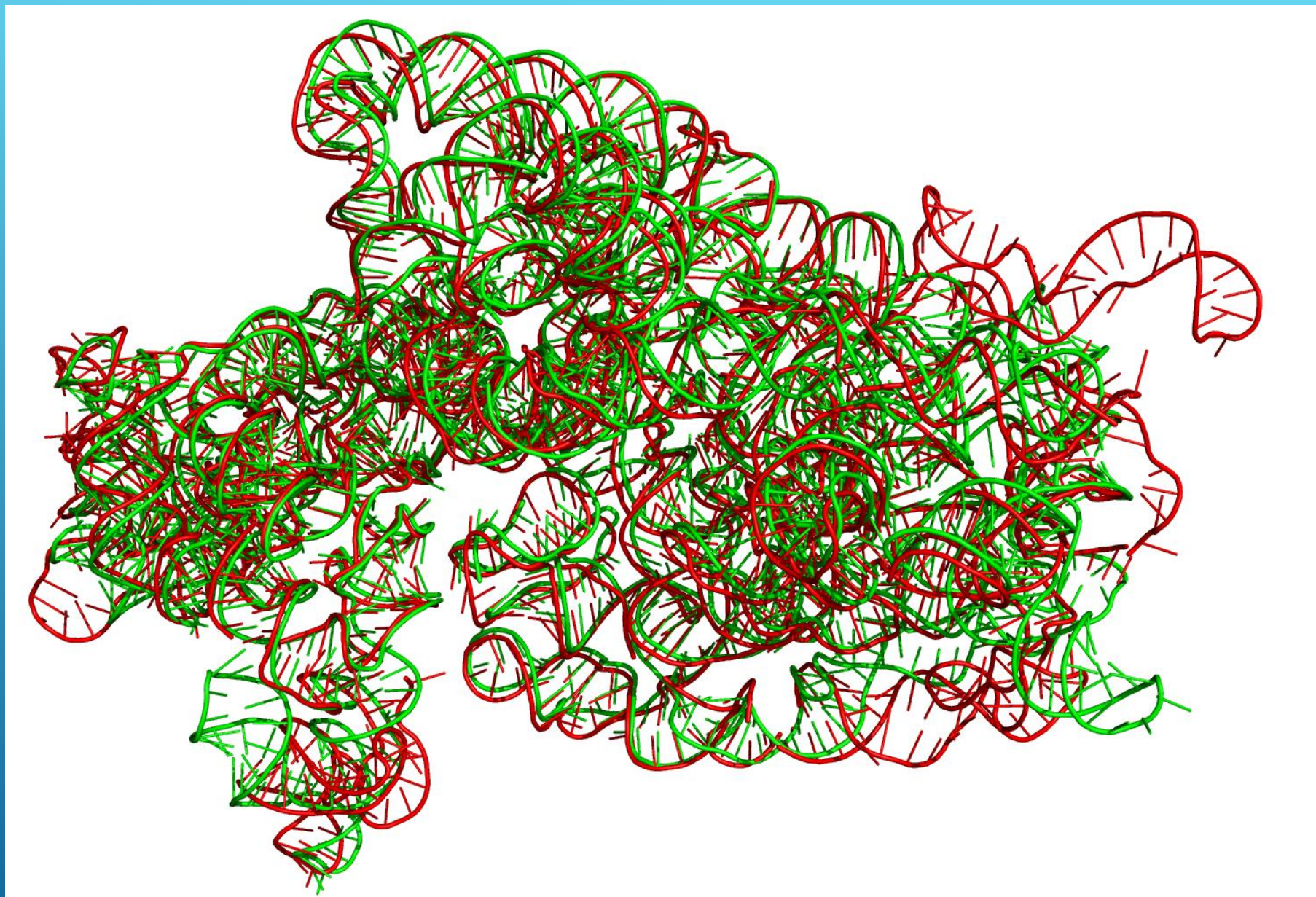
- ▶ 2QUW
- ▶ 57nt
- ▶ 80% similarity
- ▶ RMSD 4.95 vs 21.59

RESULTS

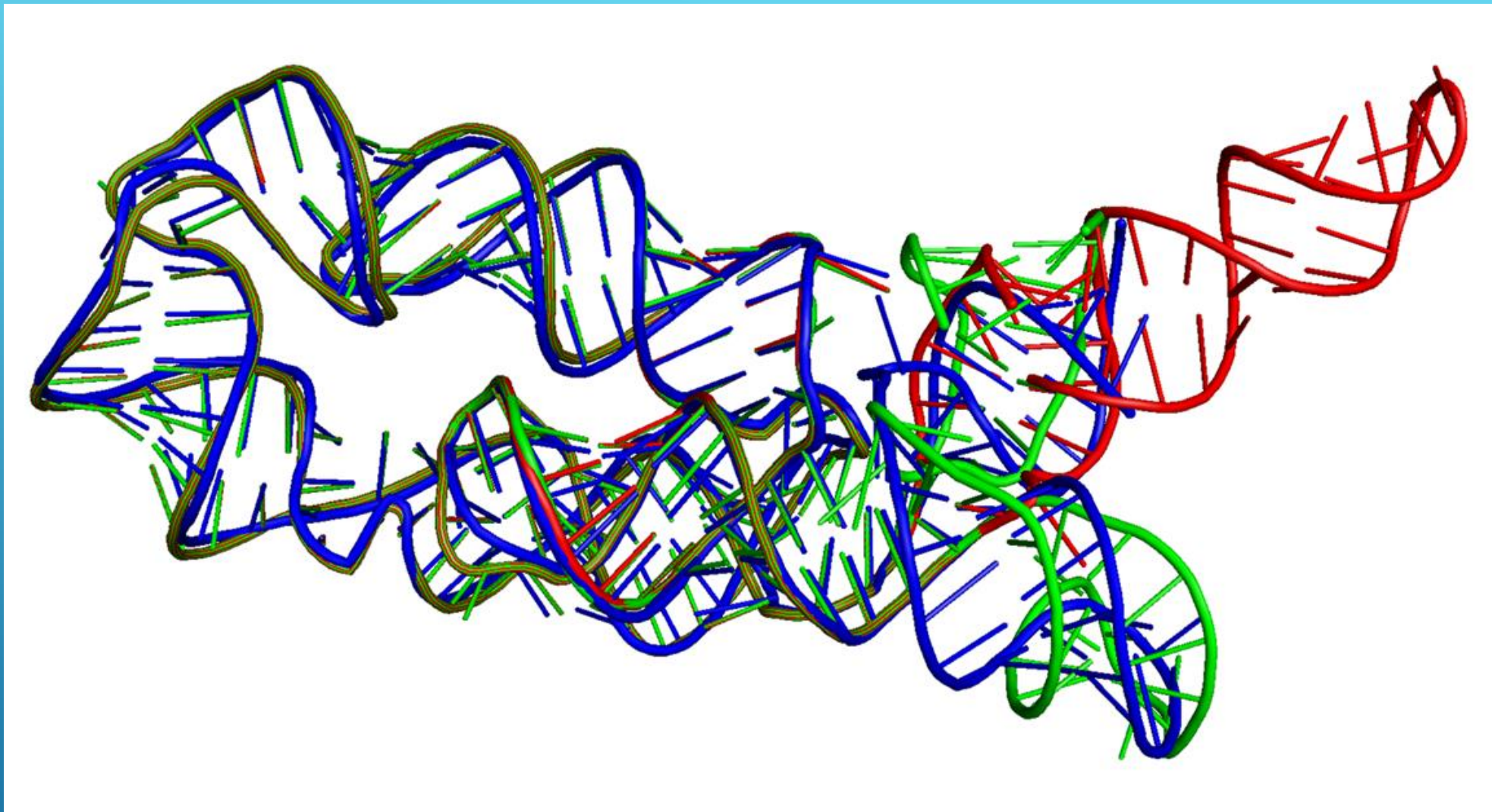


- ▶ 4IJ1
- ▶ 71.6%
- ▶ 1522 nt
- ▶ RMSD 14.5

RESULTS



- ▶ 3DIG
- ▶ 175nt



RESULTS

RESULTS

Size	Trooper without SS	Trooper with SS	ModeRNA
50 - 100	6.91	3.46	3.72
100 - 500	5.80	8.23	8.54

- ▶ Multiple templates
- ▶ Improve processing of inconsistencies in PDB and FASTA input files
- ▶ Get rid of FARFAR

FUTURE WORK?

A decorative graphic consisting of several parallel white lines of varying lengths, slanted upwards from left to right, located in the bottom right corner of the slide.