



# When will RNA get its AlphaFold Moment?

Bohdan Schneider, et al.

Stephen Owsney



# Overview

RNA Refresher

Motivation

RNA Background

RNA 3D Structure Prediction

RNA-Puzzles / CASP-RNA

Challenges

Conclusions / Future Work



# RNA Background

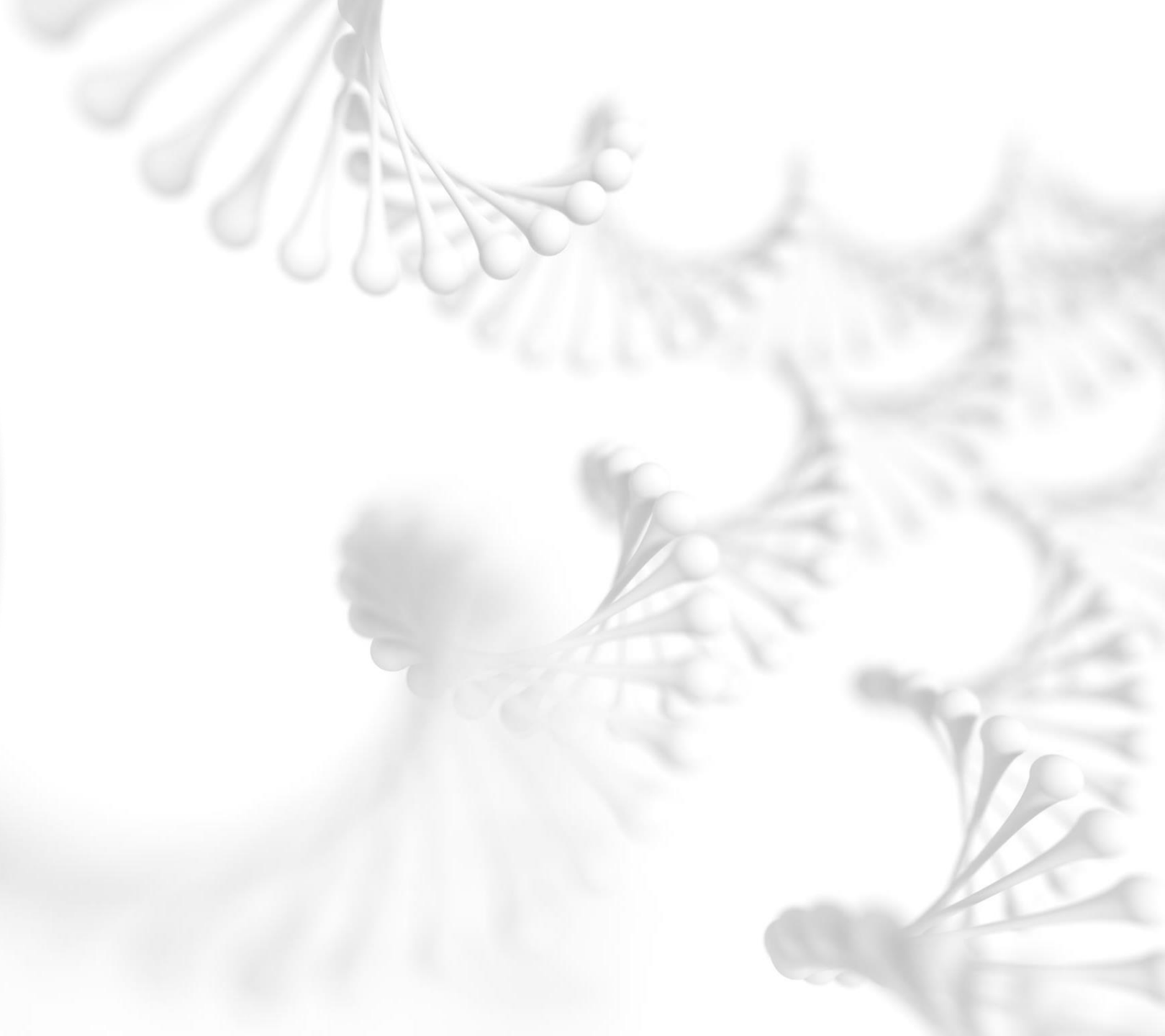
---



## RNA Background

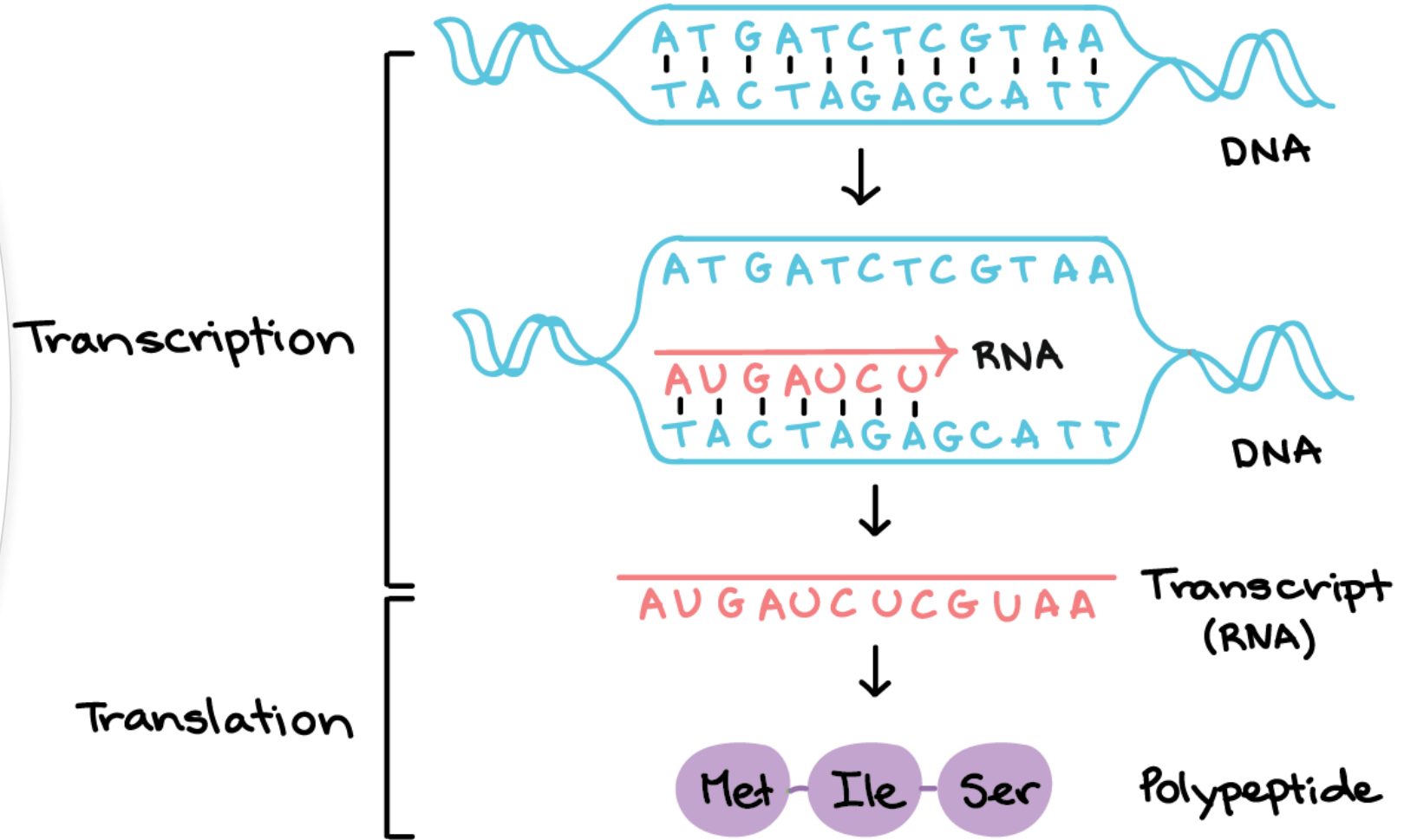
---

- Messenger RNA (mRNA)



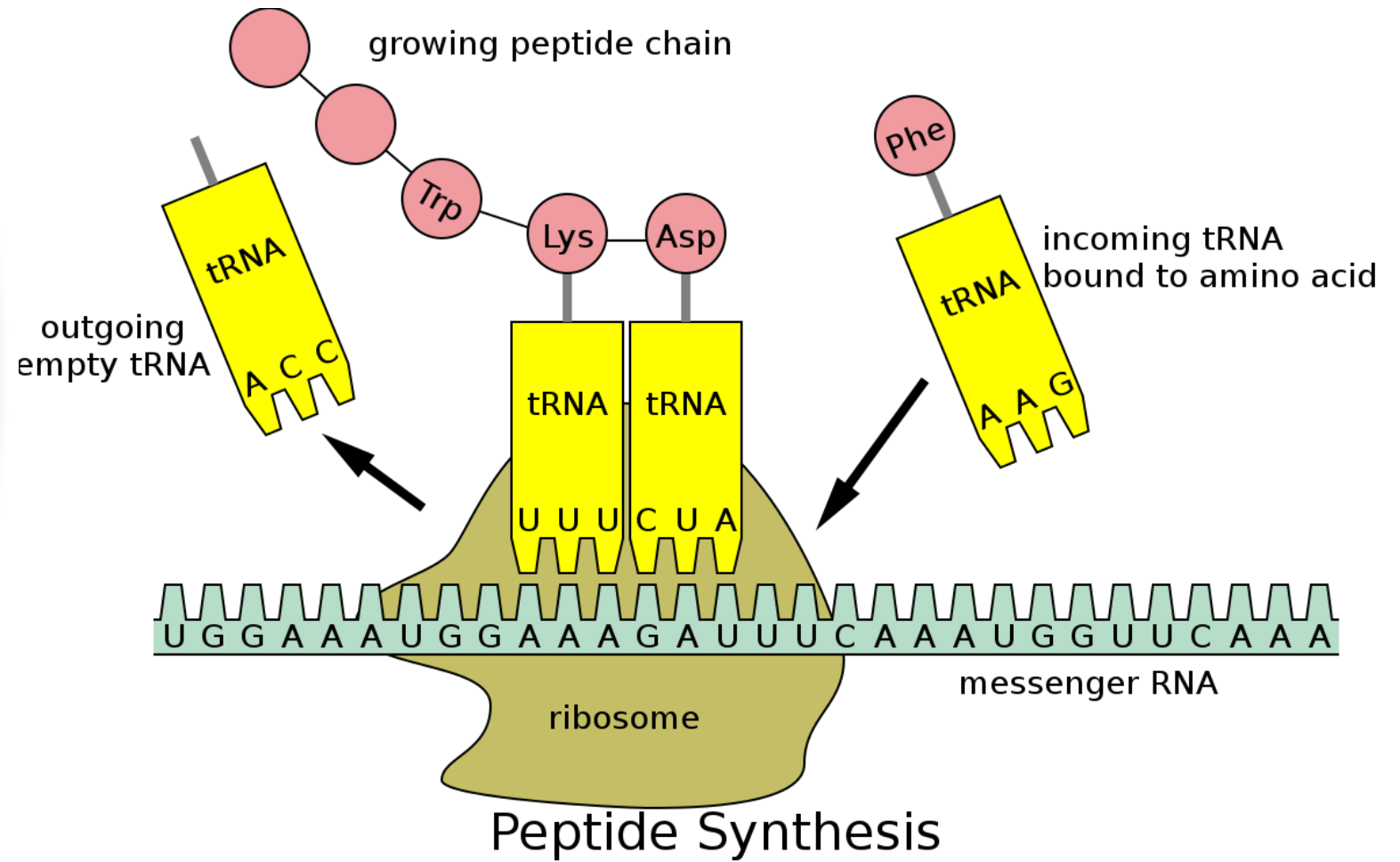
# RNA Background

- Messenger RNA (mRNA)



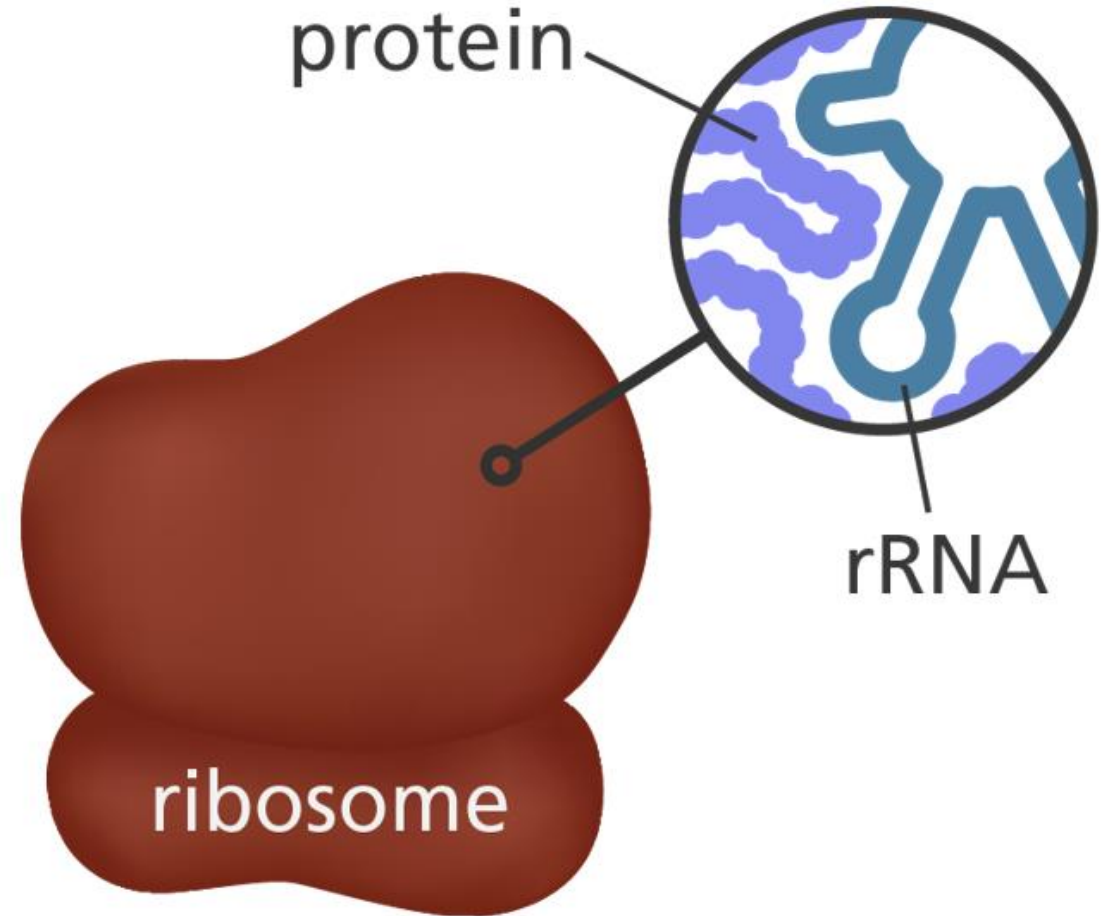
# RNA Background

- Messenger RNA (mRNA)
- Transfer RNA (tRNA)



## RNA Background

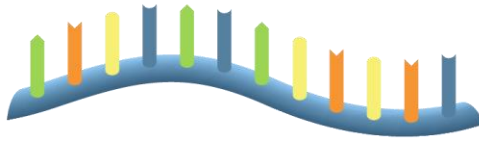
- Messenger RNA (mRNA)
- Transfer RNA (tRNA)
- Ribosomal RNA (rRNA)



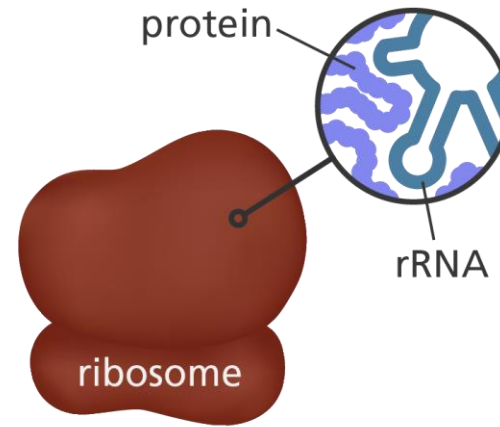
**ribosomal RNA  
(rRNA)**

## RNA Background

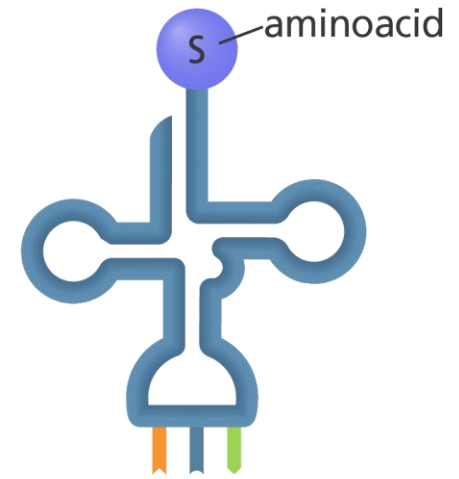
- Messenger RNA (mRNA)
- Transfer RNA (tRNA)
- Ribosomal RNA (rRNA)



**messenger RNA  
(mRNA)**



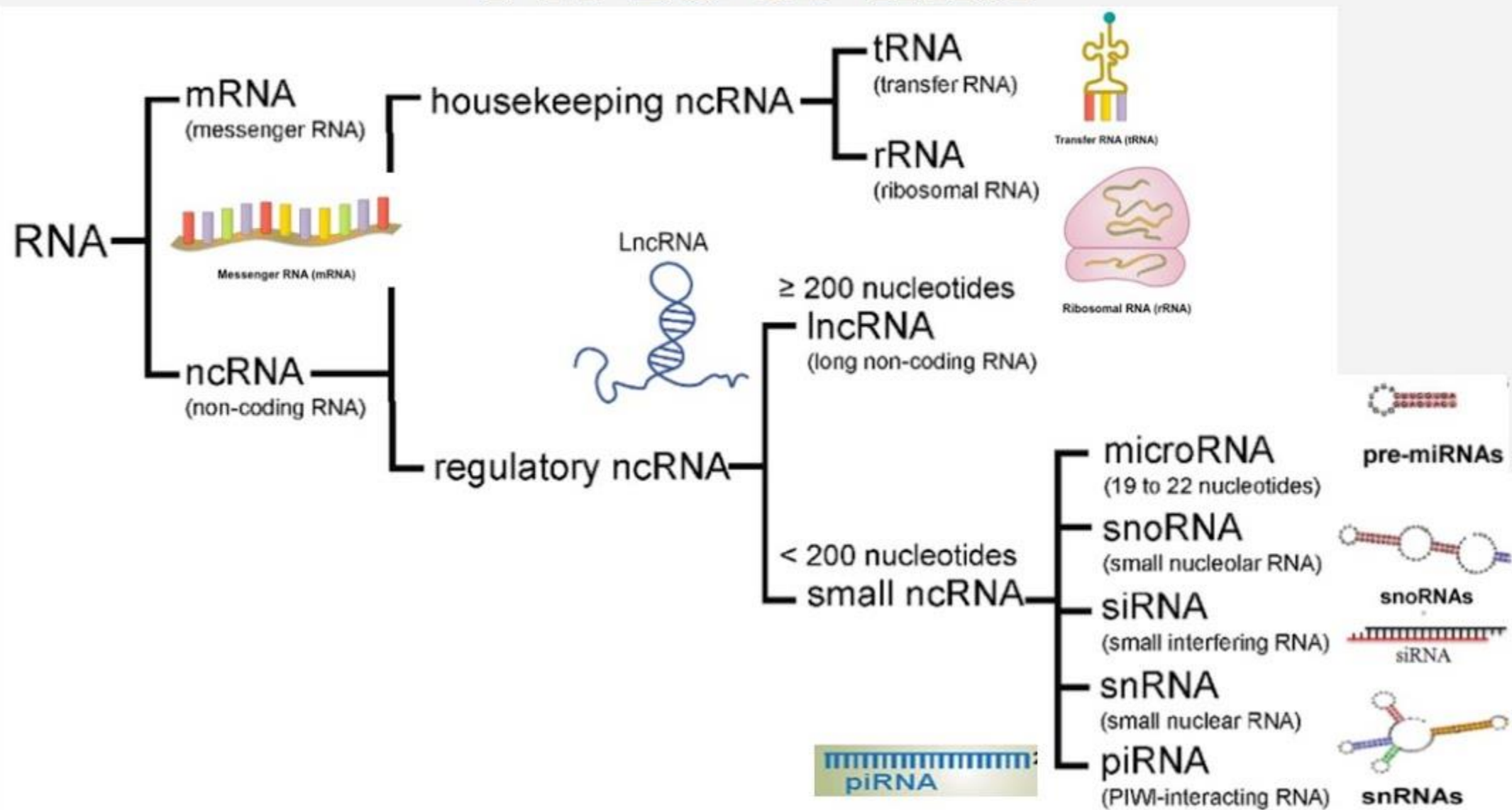
**ribosomal RNA  
(rRNA)**



**transfer RNA  
(tRNA)**



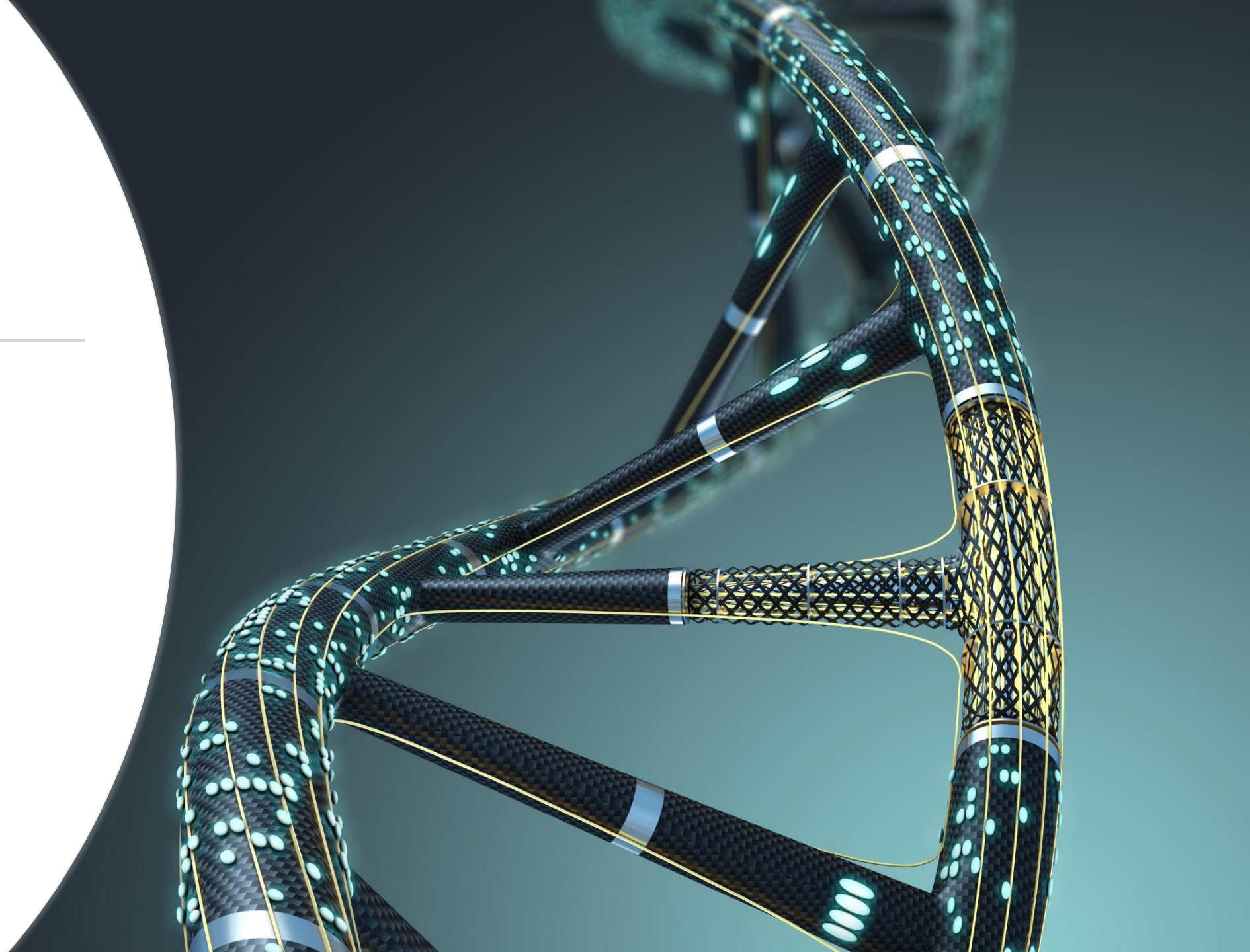
# TYPES OF RNA





# Motivation

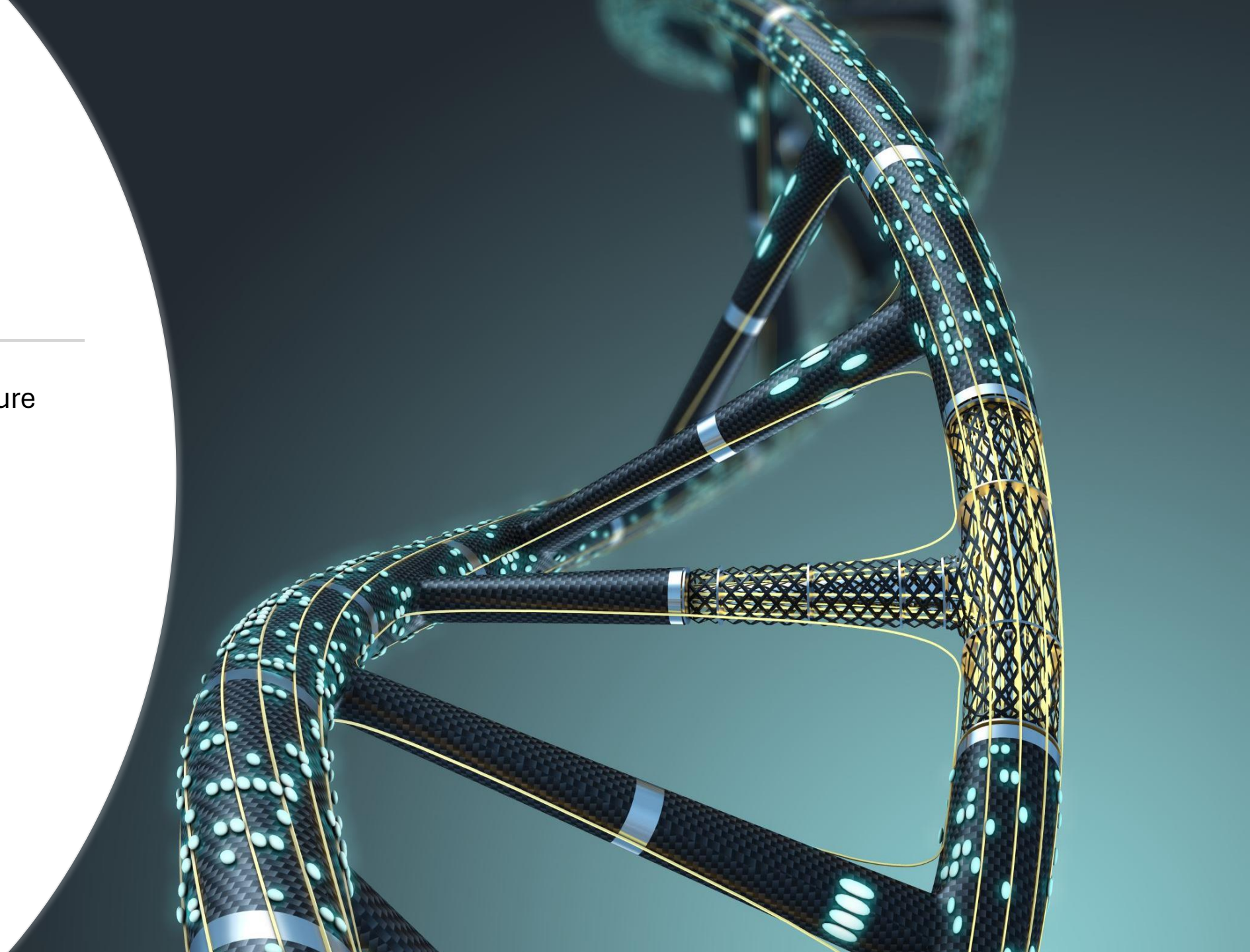
---



# Motivation

---

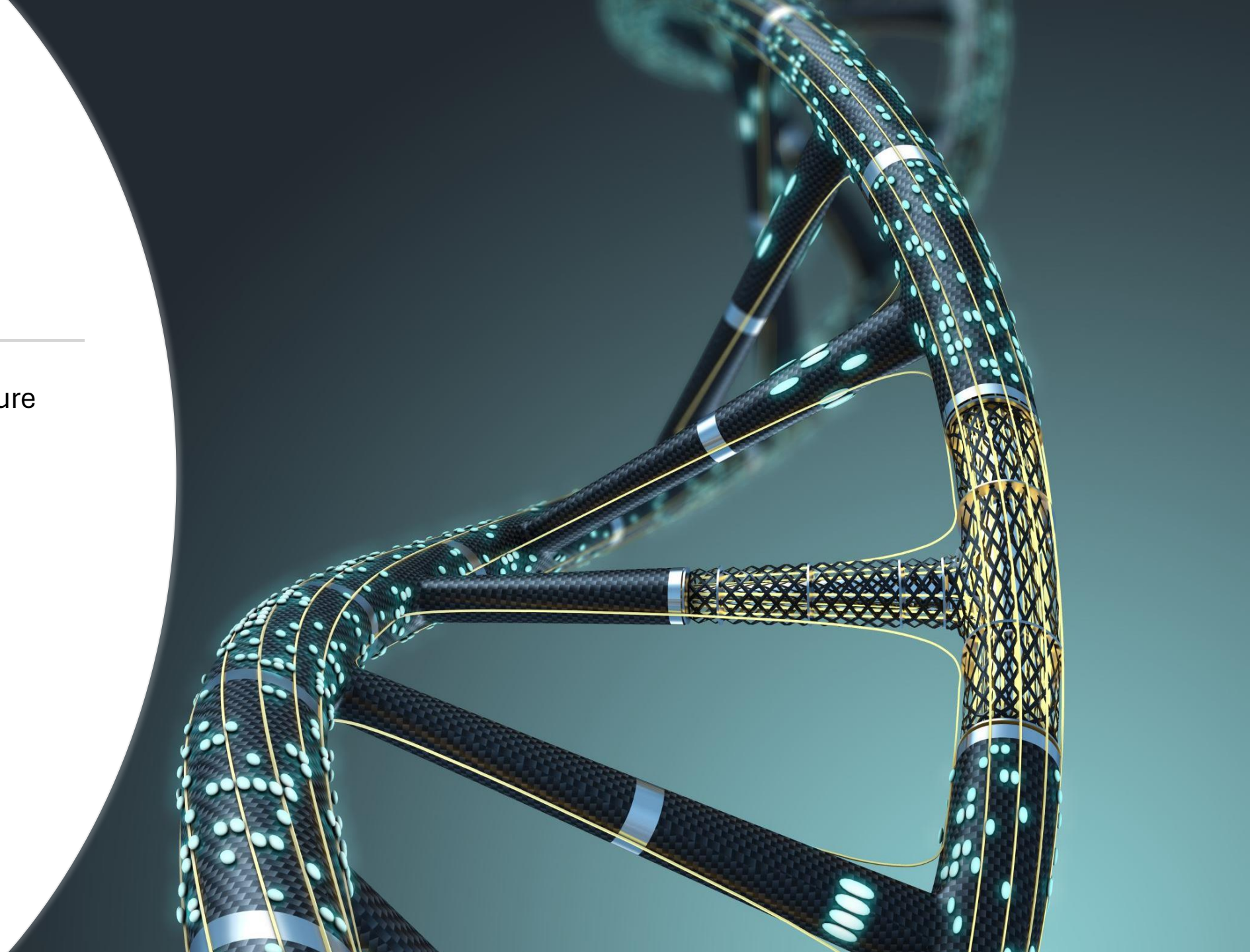
- RNA function tied to structure



# Motivation

---

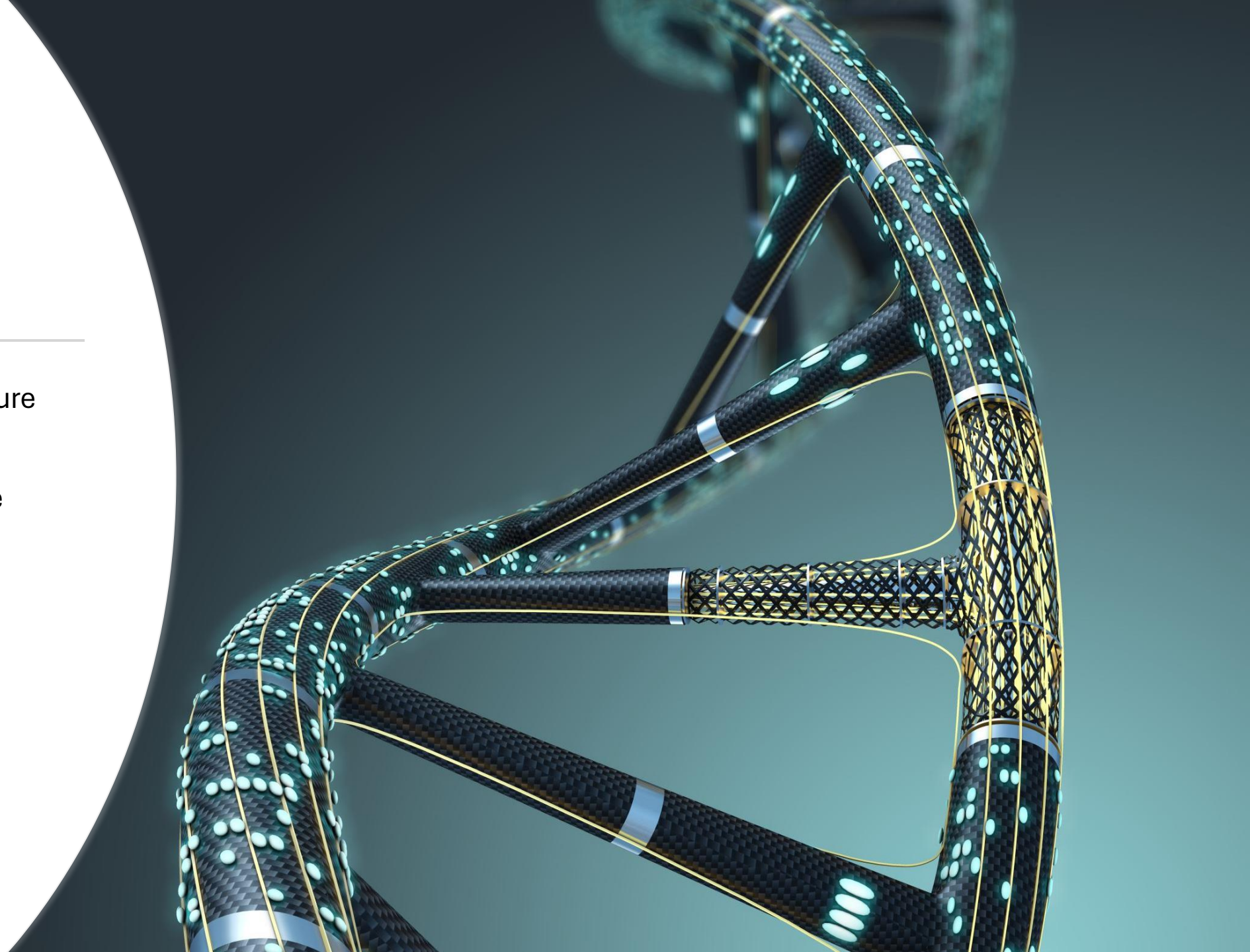
- RNA function tied to structure
- Advancements in science



# Motivation

---

- RNA function tied to structure
- Advancements in science
- Advancements in medicine

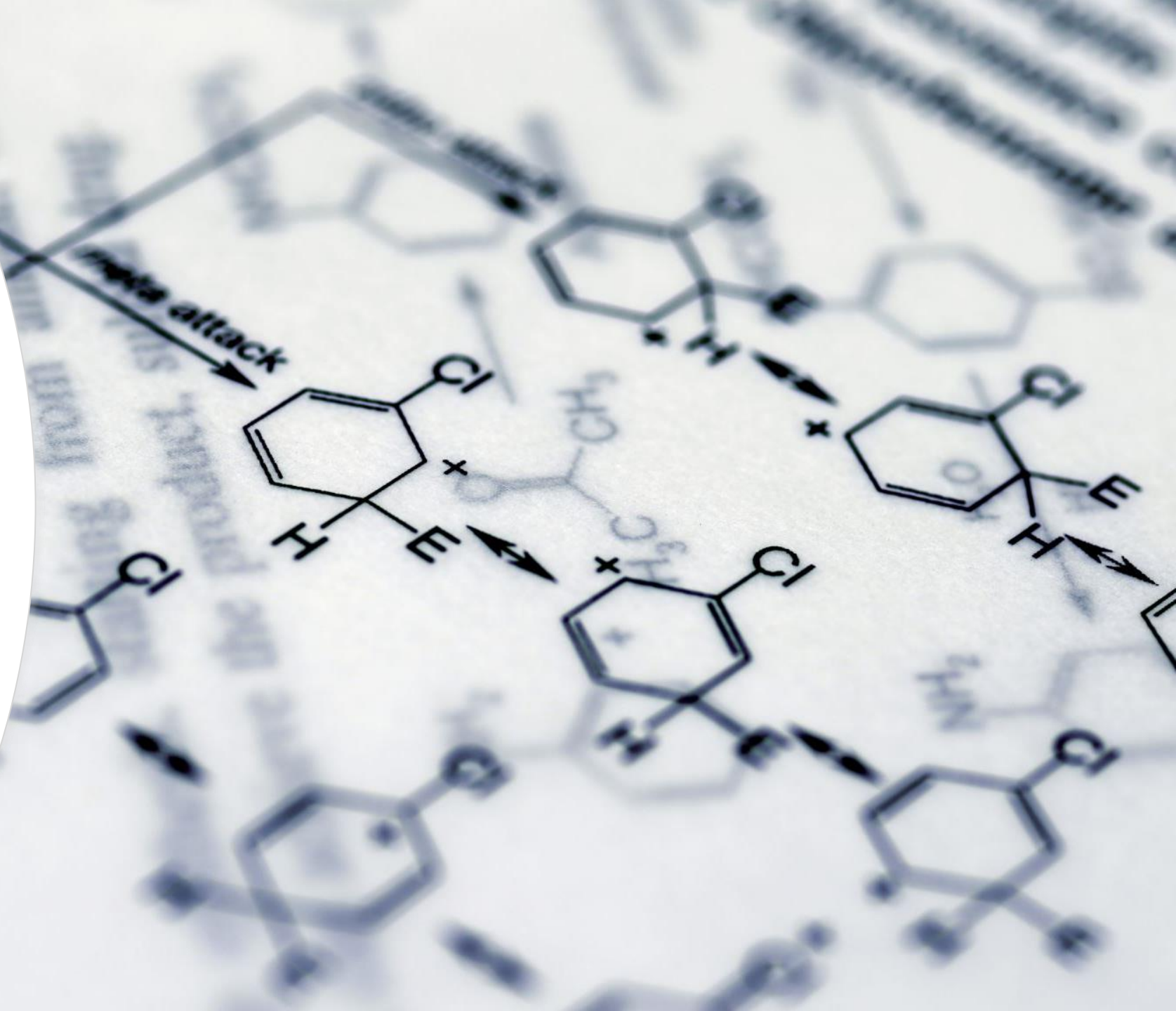




## RNA: A closer look

---

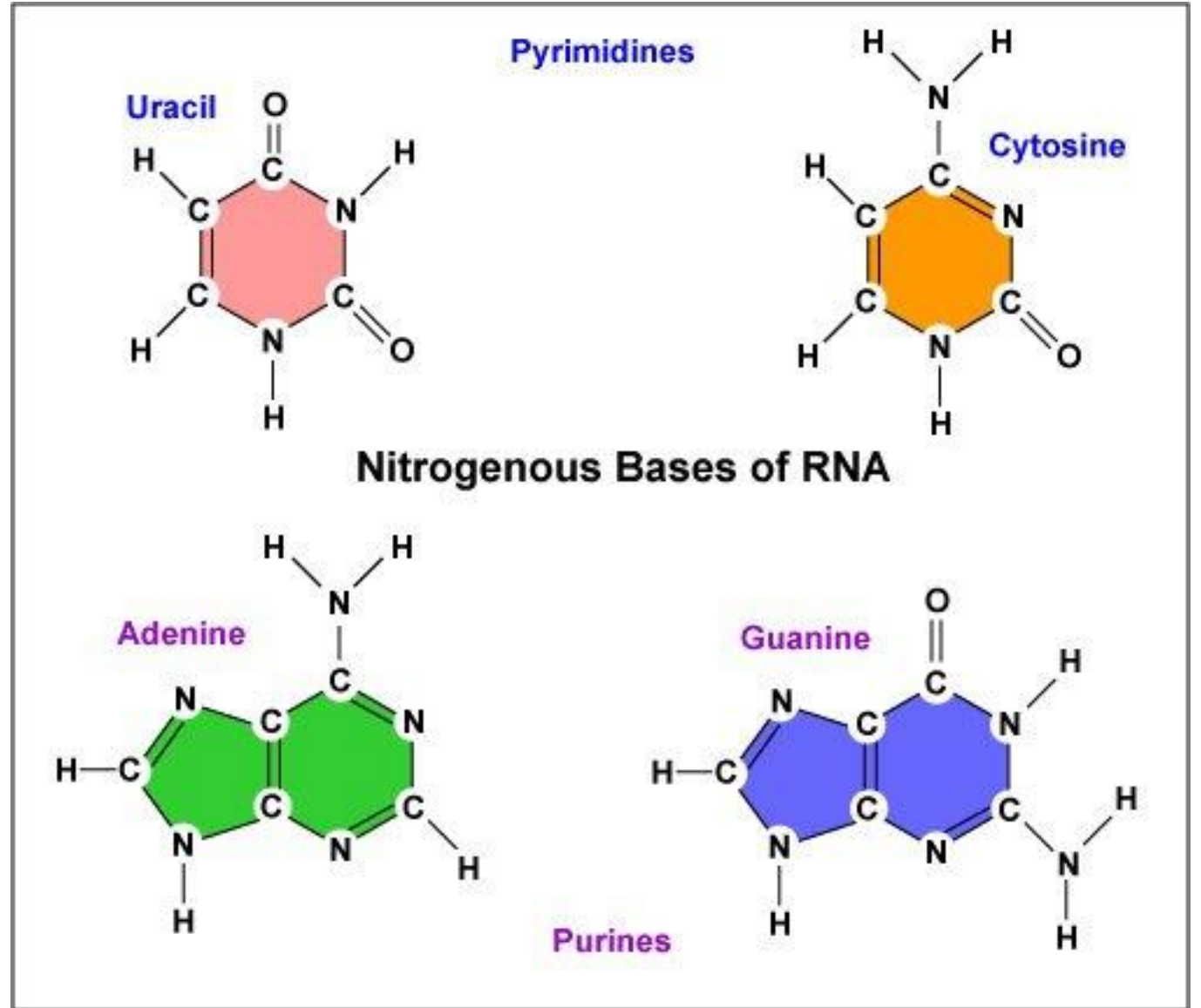
Nucleotides break down into three key elements



# RNA: A closer look

Nucleotides break down into three key elements

- Aromatic nitrogenous bases

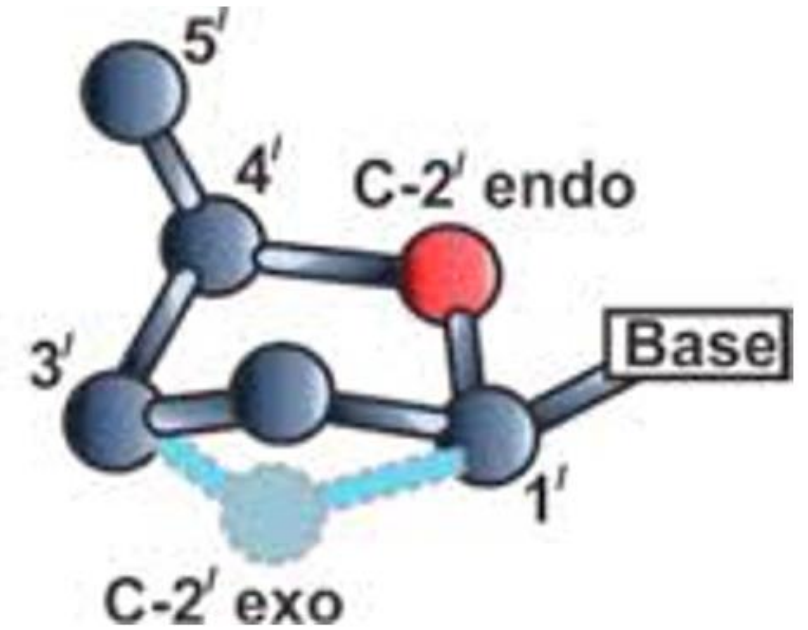
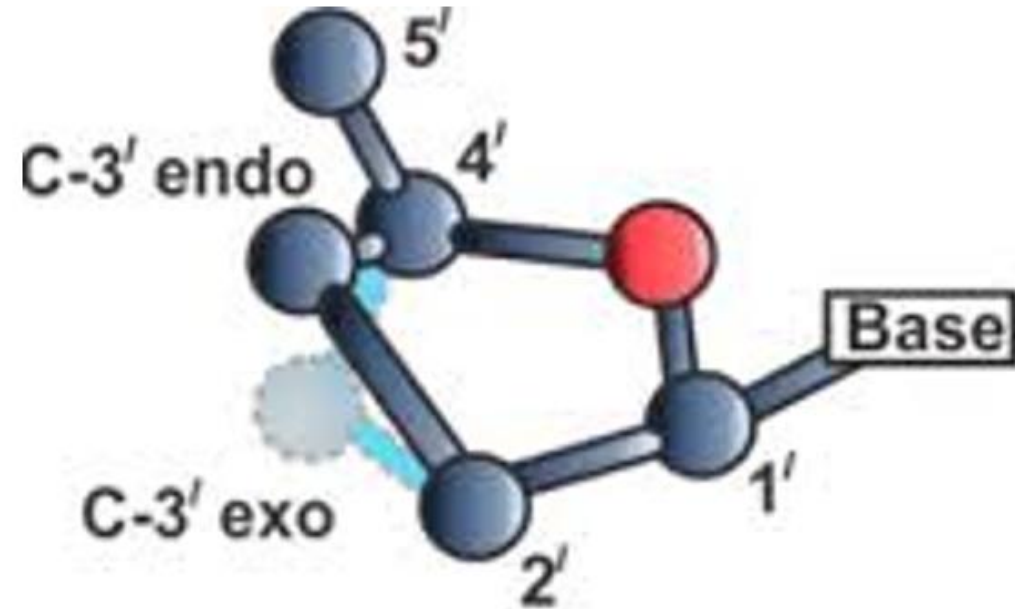




# RNA: A closer look

Nucleotides break down into three key elements

- Aromatic nitrogenous bases
- Ribose sugar rings

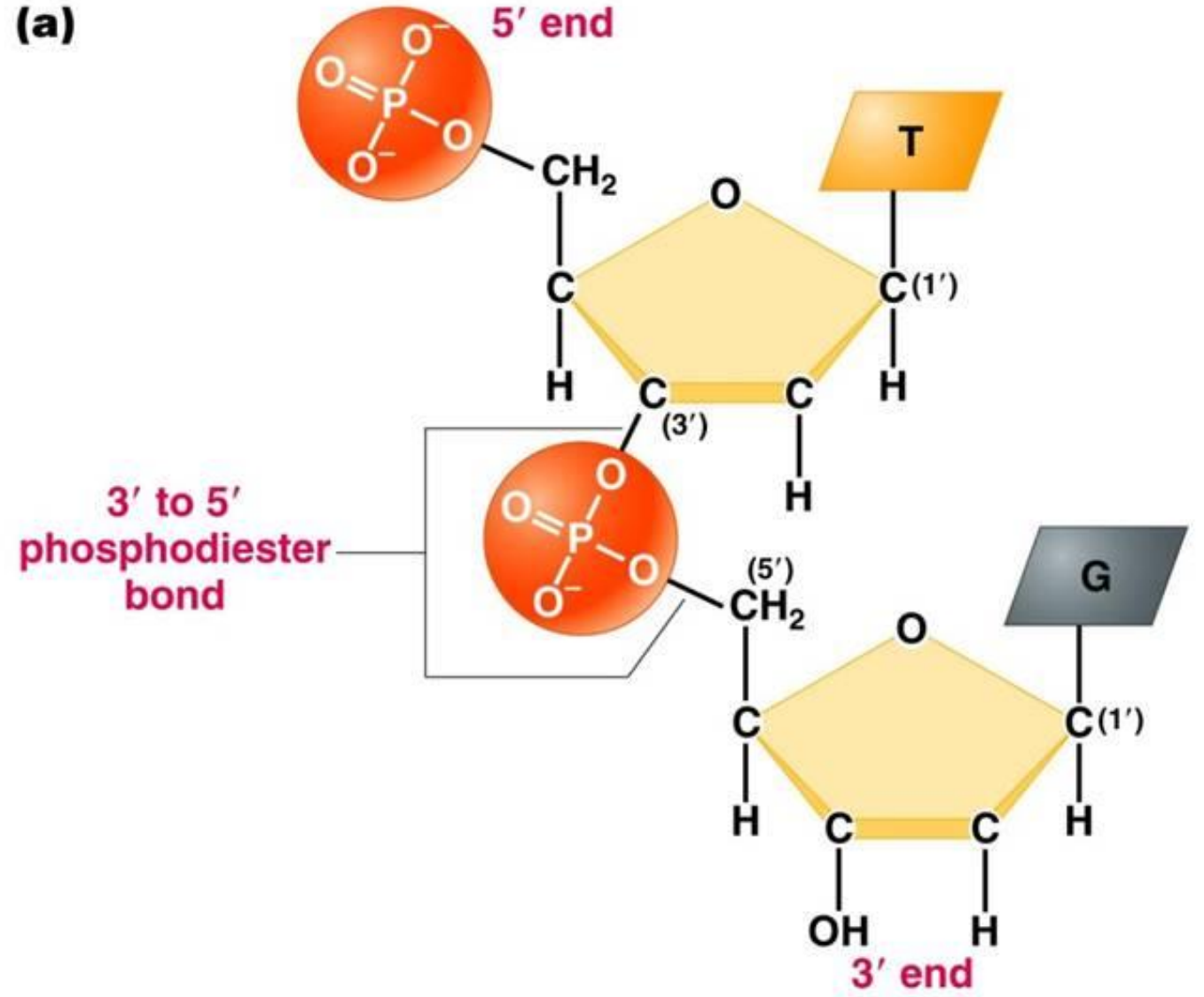


# RNA: A closer look

Nucleotides break down into three key elements

- Aromatic nitrogenous bases
- Ribose sugar rings
- Phosphate groups

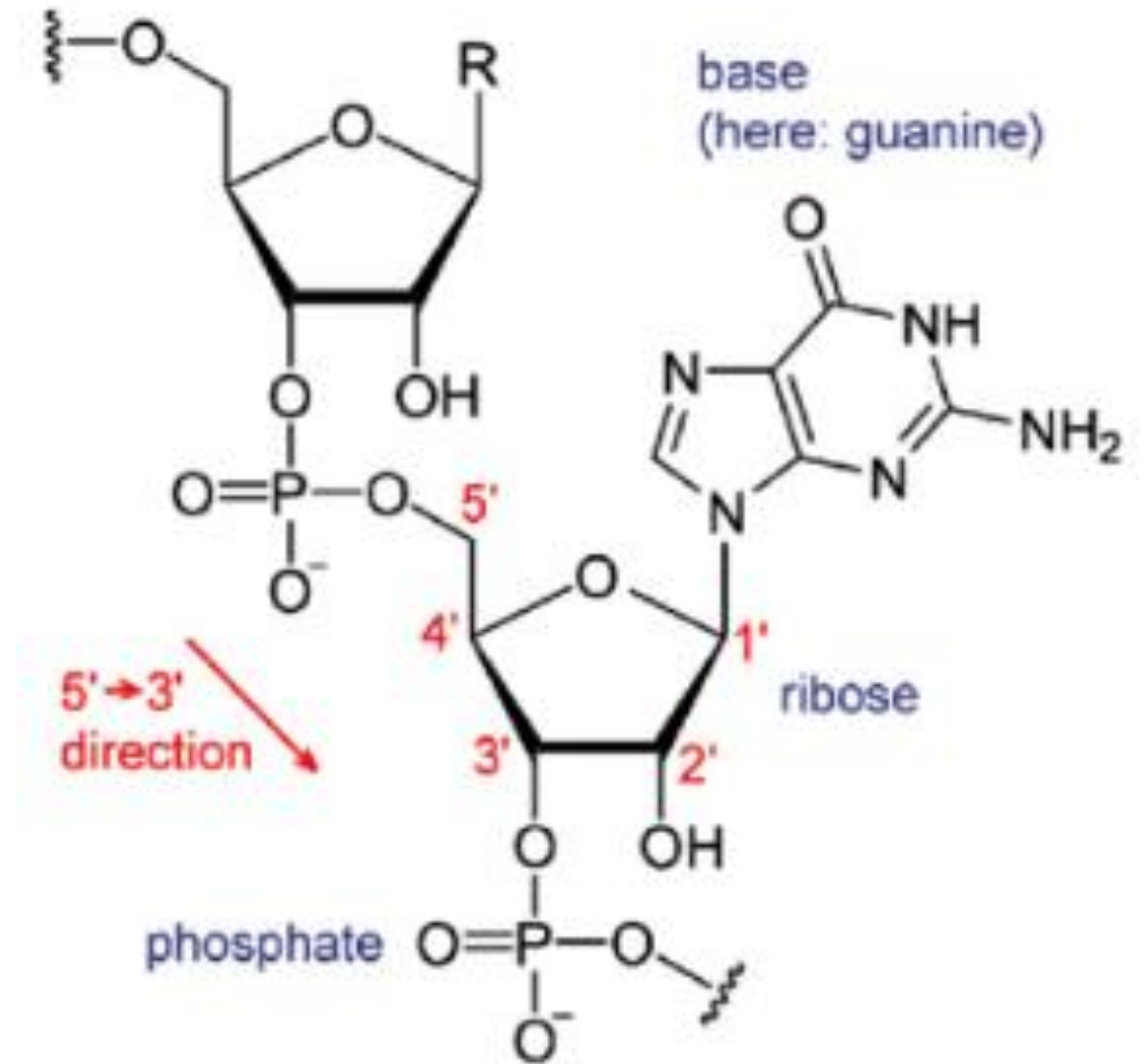
(a)



# RNA: A closer look

Nucleotides break down into three key elements

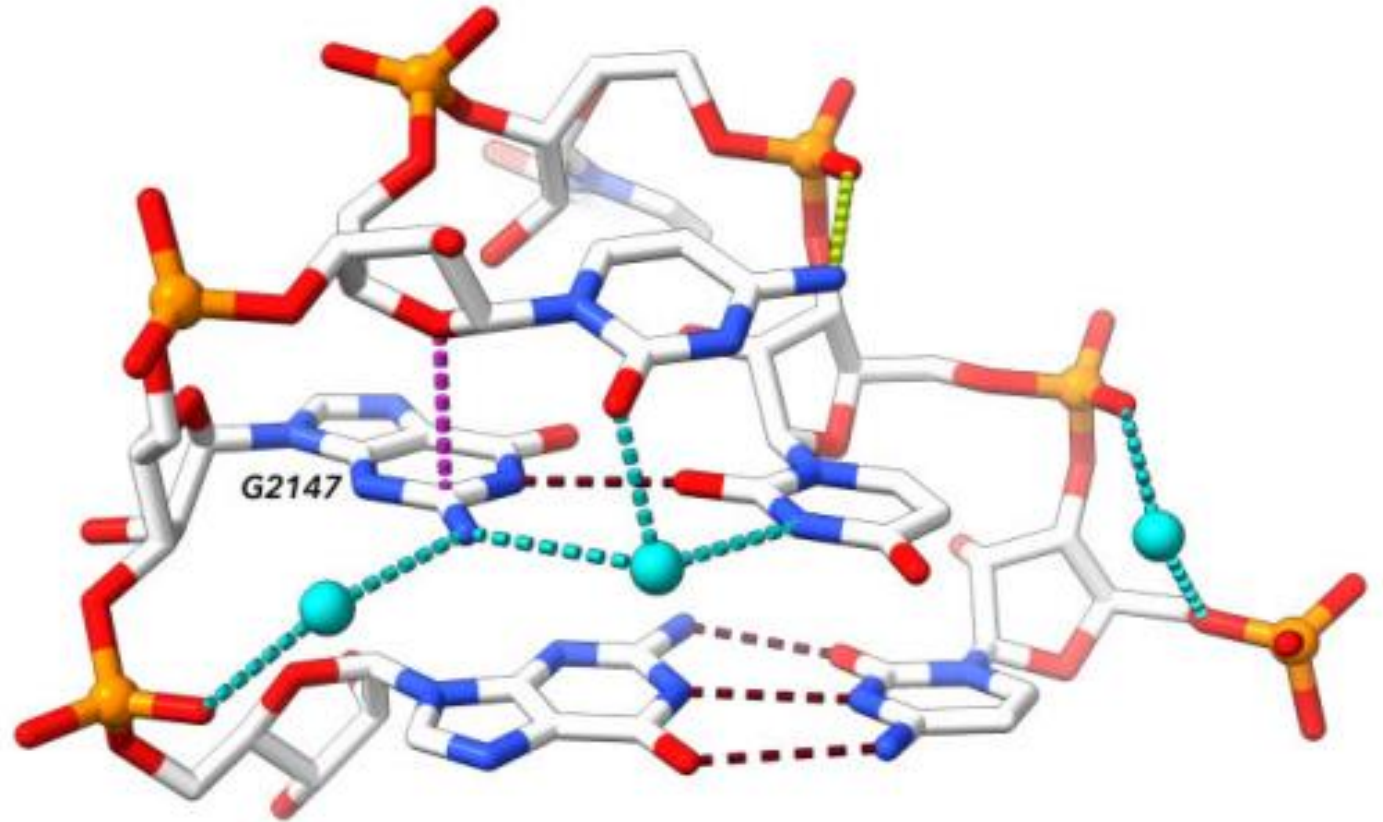
- Aromatic nitrogenous bases
- Ribose sugar rings
- Phosphate groups



# RNA: A closer look

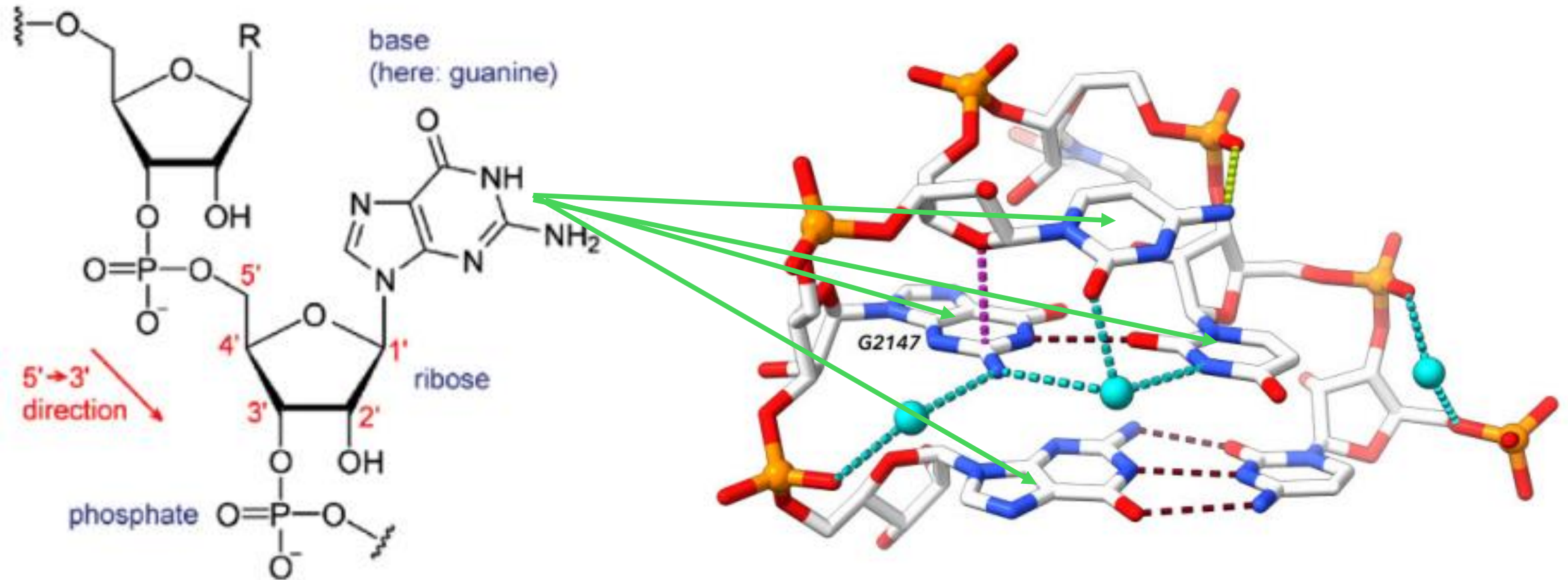
Nucleotides break down into three key elements

- Aromatic nitrogenous bases
- Ribose sugar rings
- Phosphate groups



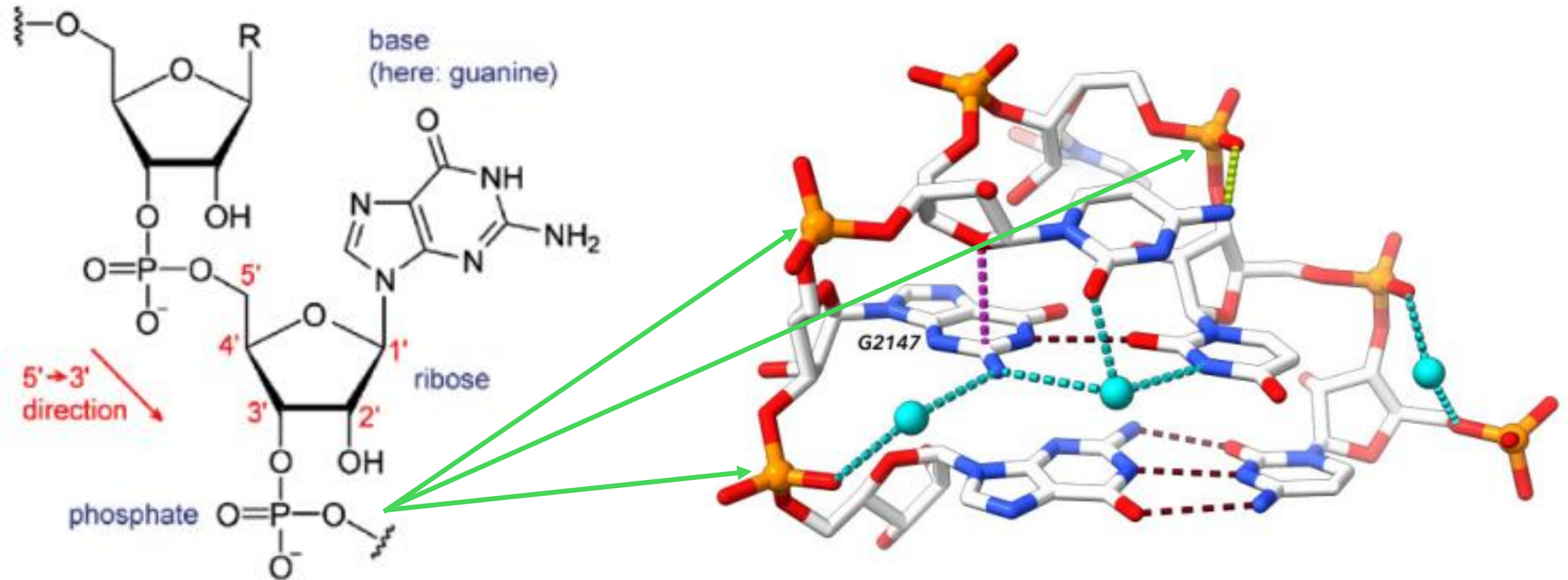
# RNA: A closer look

---



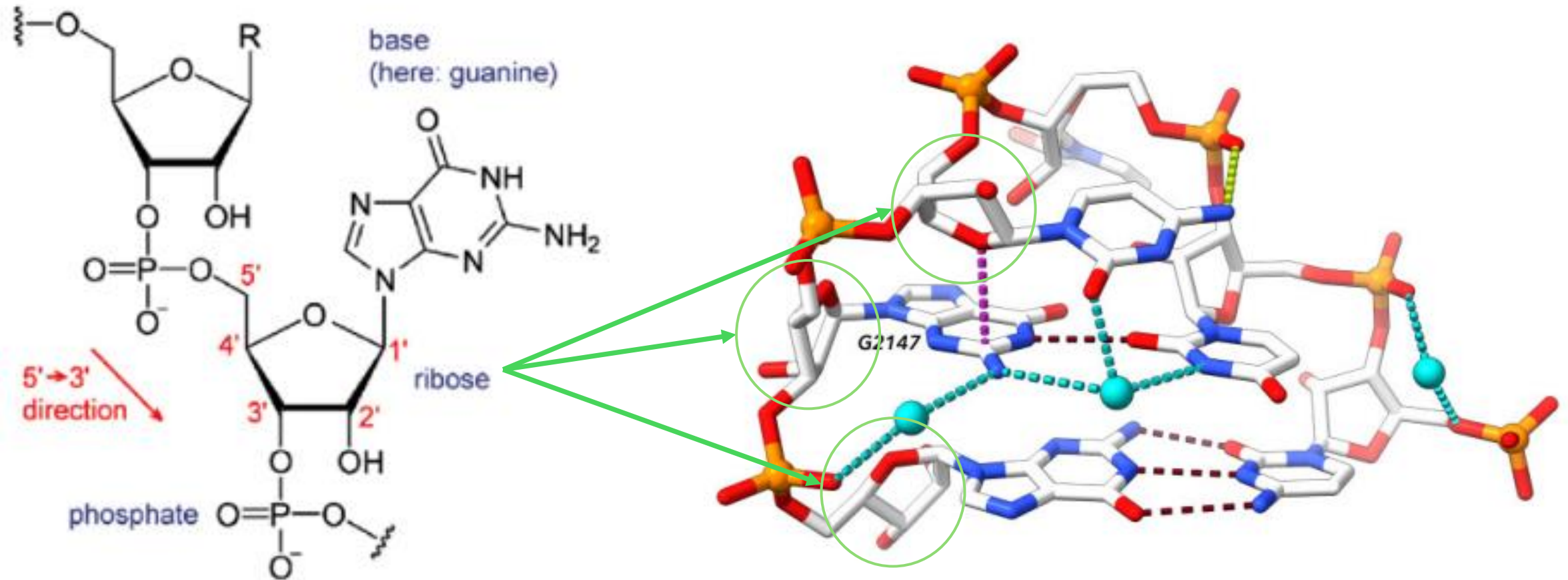
# RNA: A closer look

---



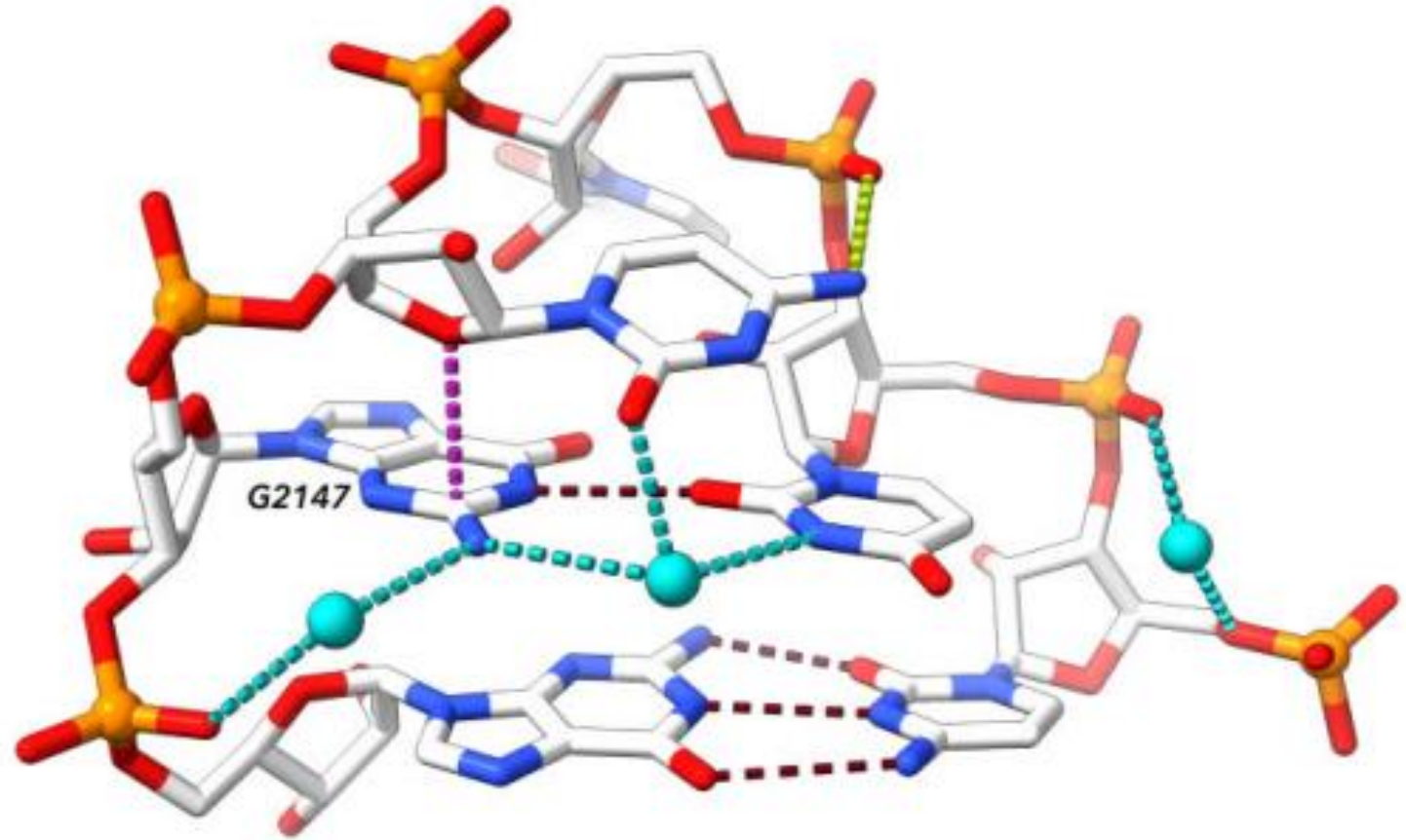
# RNA: A closer look

---



## RNA: A closer look

RNA is structurally determined by interactions

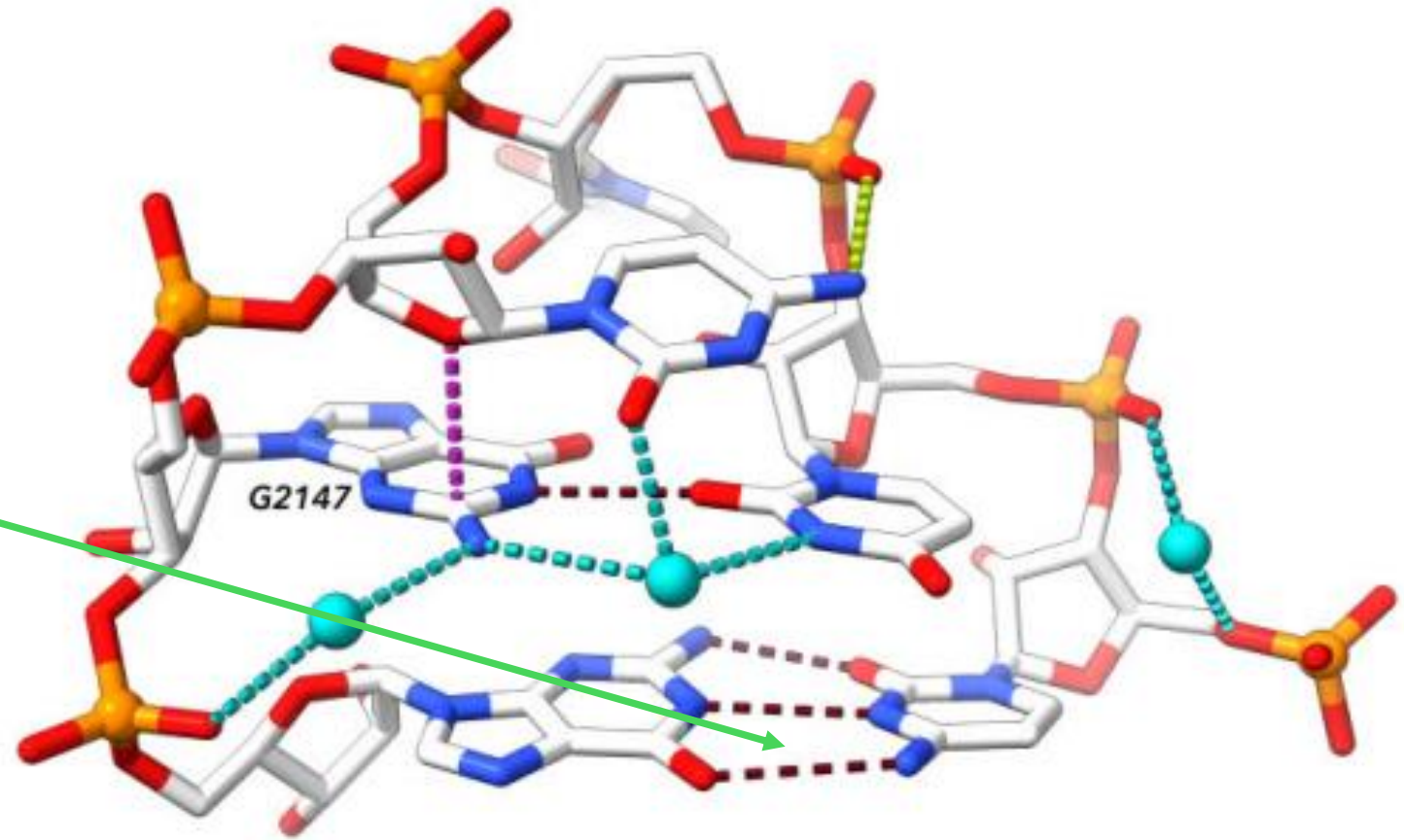




## RNA: A closer look

RNA is structurally determined by interactions

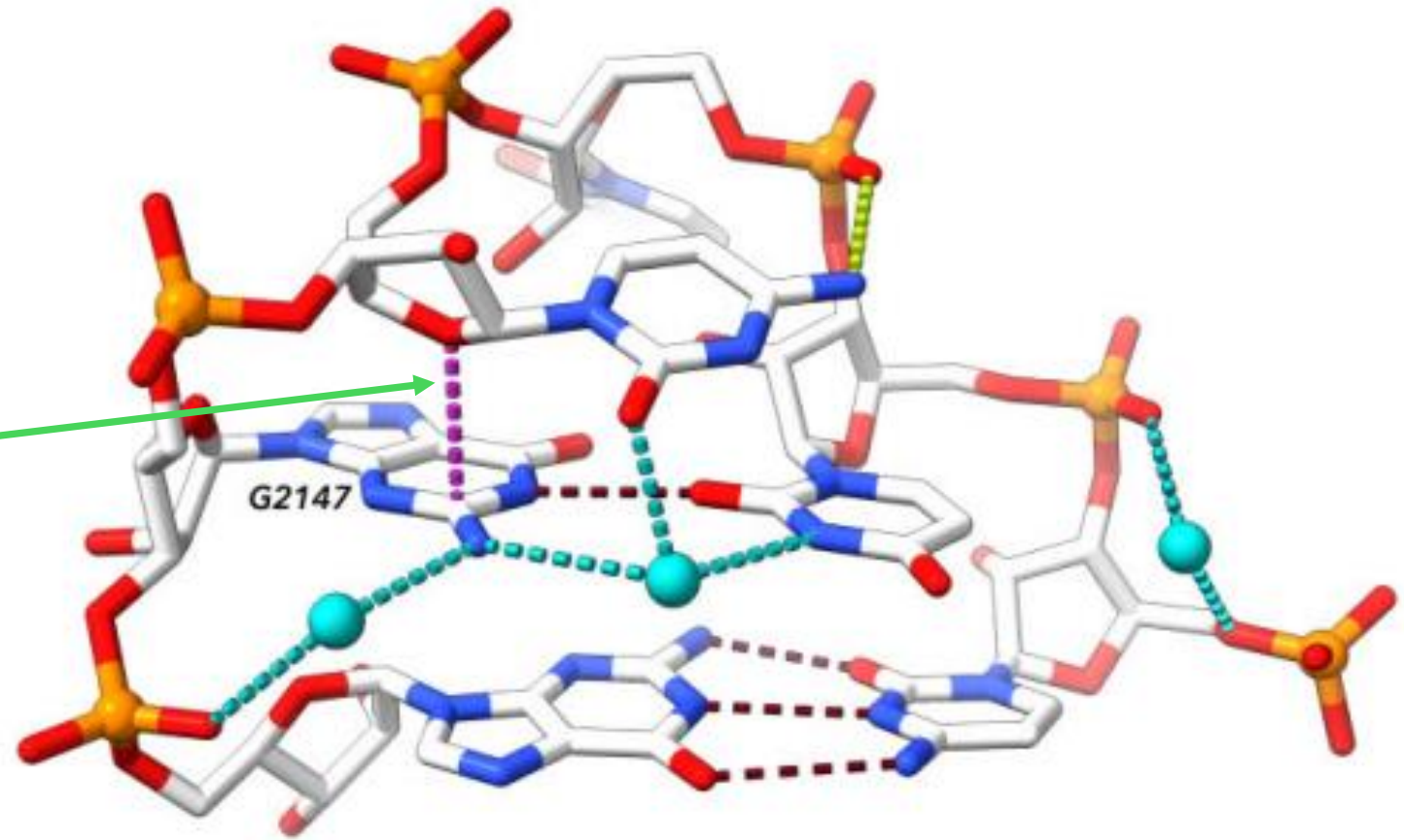
- Base pairing hydrogen bonds



## RNA: A closer look

RNA is structurally determined by interactions

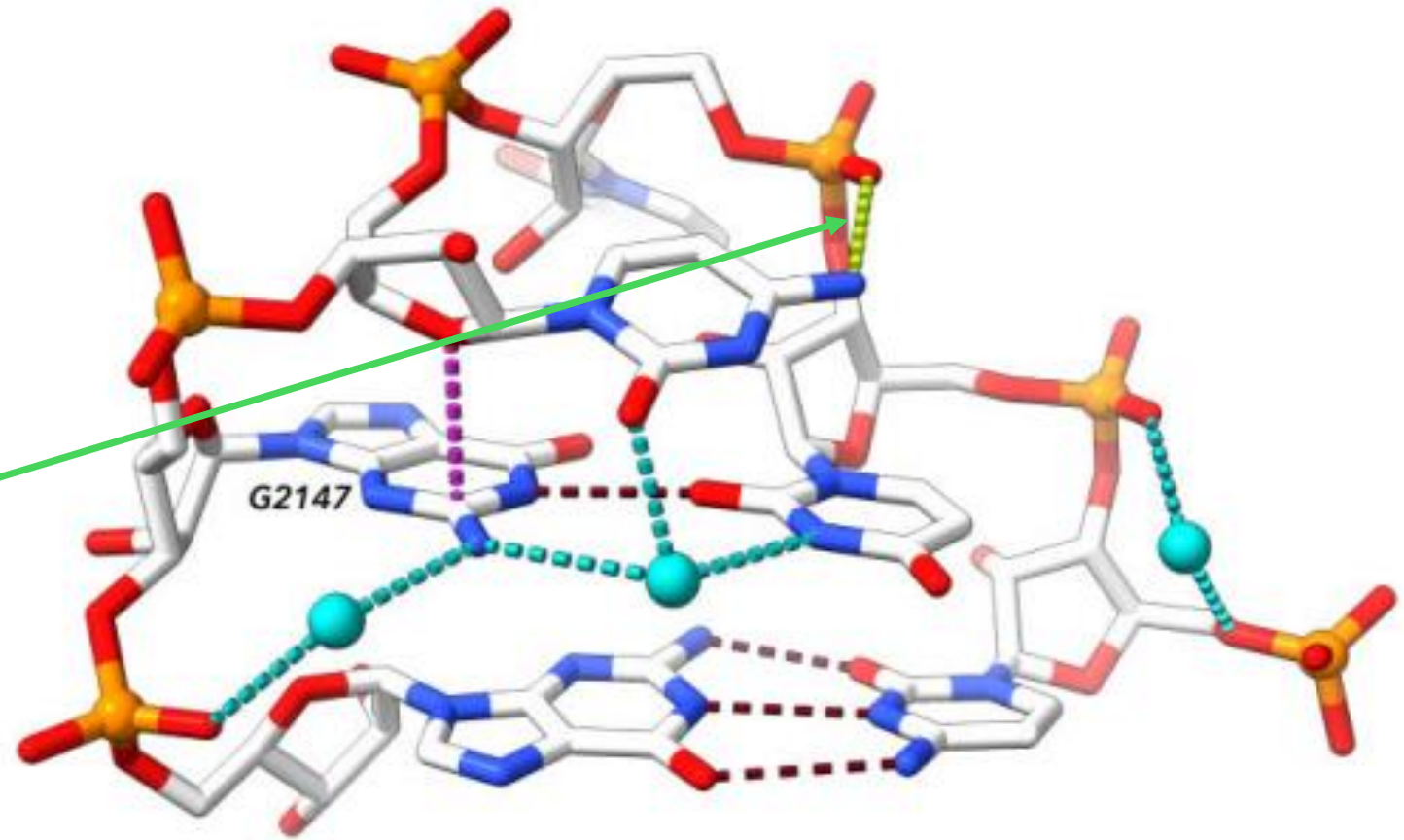
- Base pairing hydrogen bonds
- Sugar-base stacking



## RNA: A closer look

RNA is structurally determined by interactions

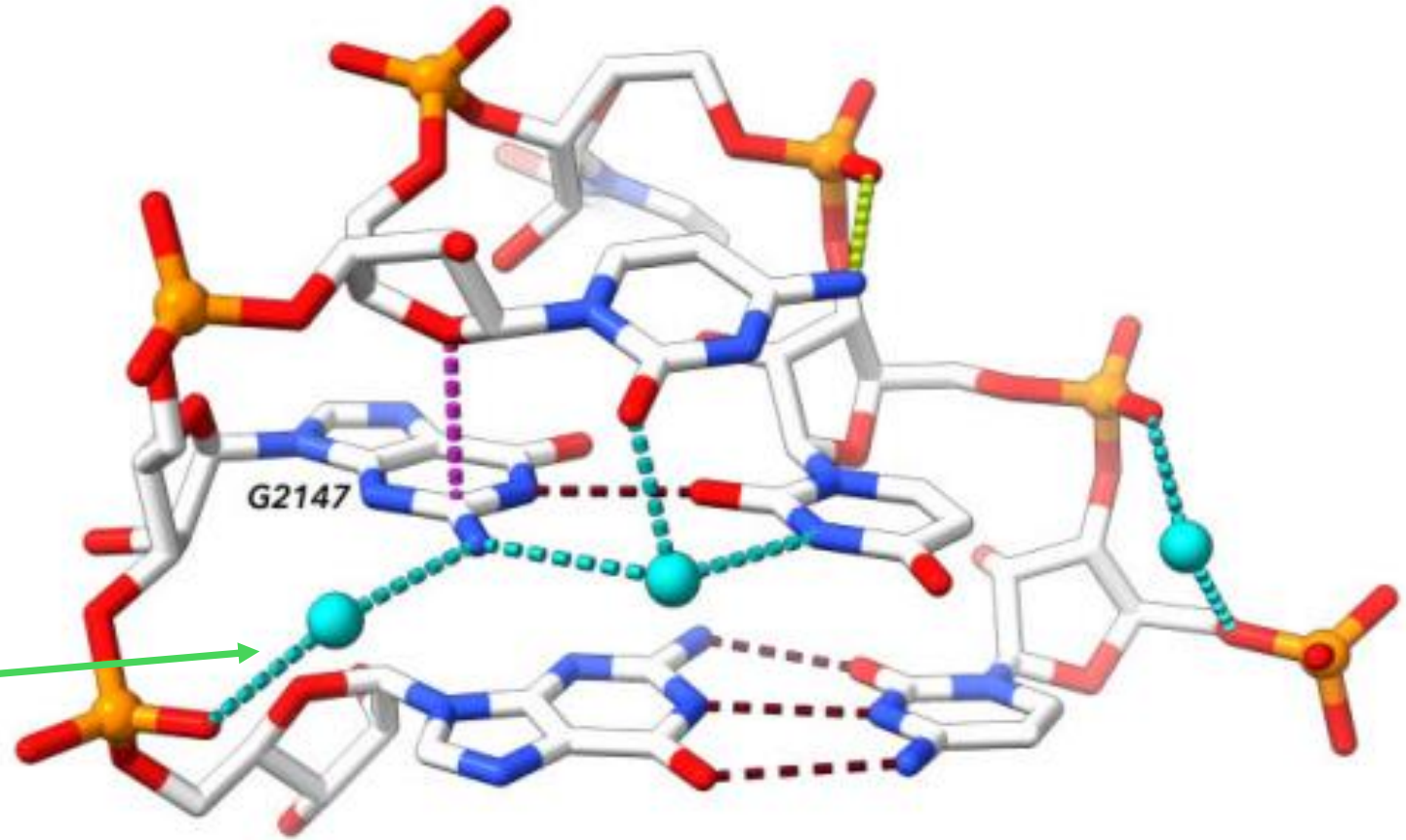
- Base pairing hydrogen bonds
- Sugar-base stacking
- Phosphate-base hydrogen bonds



## RNA: A closer look

RNA is structurally determined by interactions

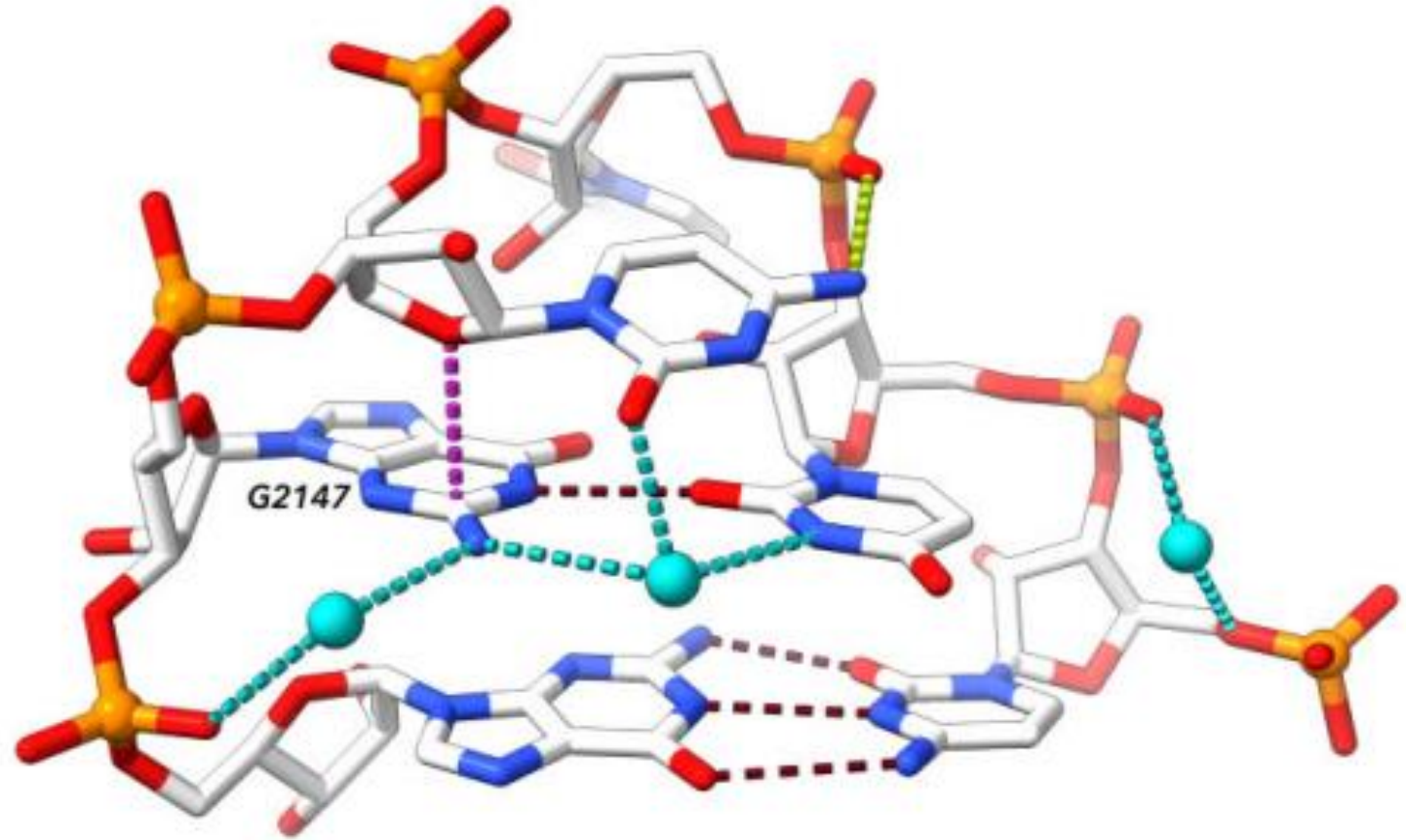
- Base pairing hydrogen bonds
- Sugar-base stacking
- Phosphate-base hydrogen bonds
- Water-formed hydrogen bonds



## RNA: A closer look

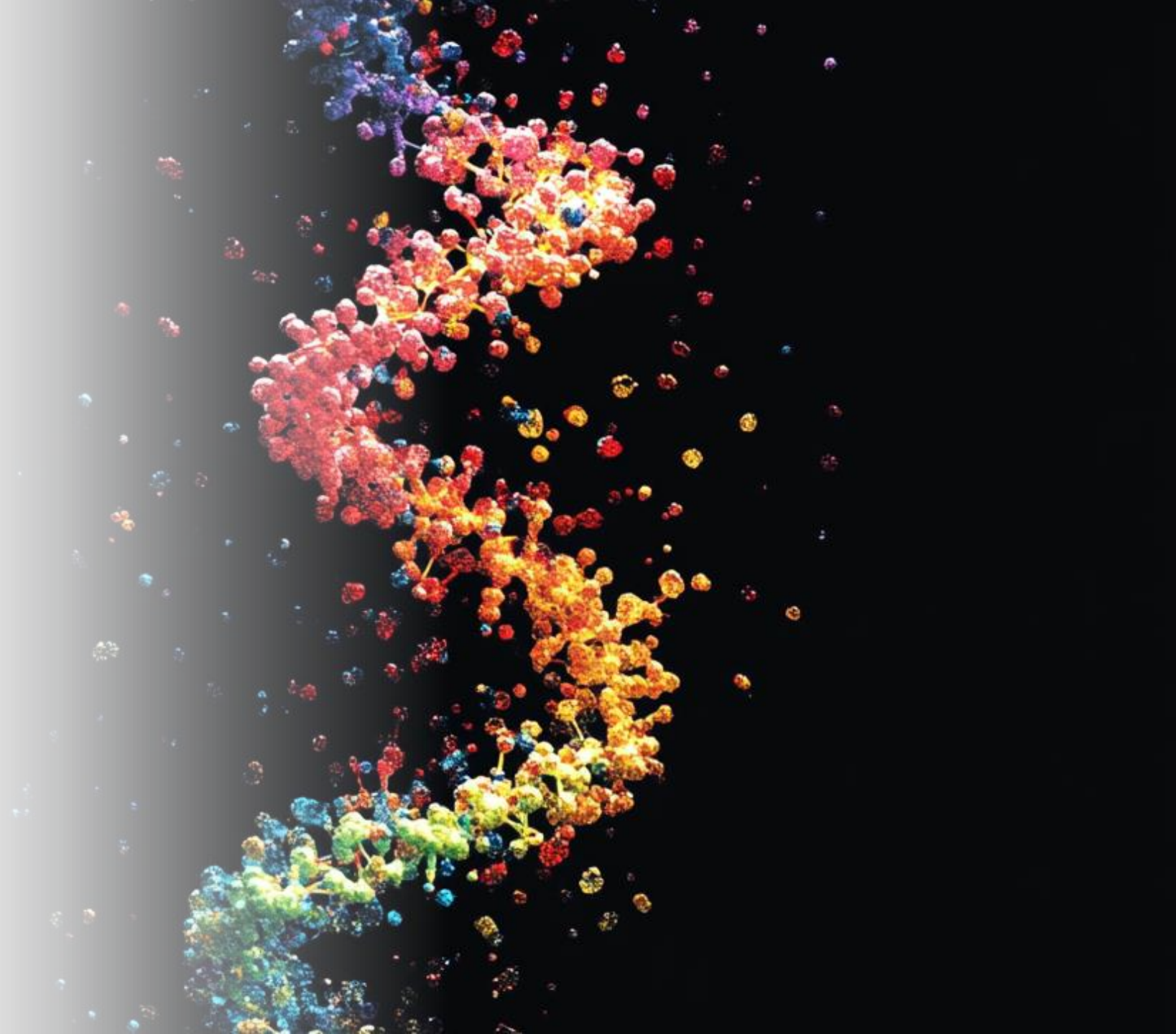
RNA is structurally determined by interactions

- Base pairing hydrogen bonds
- Sugar-base stacking
- Phosphate-base hydrogen bonds
- Water-formed hydrogen bonds





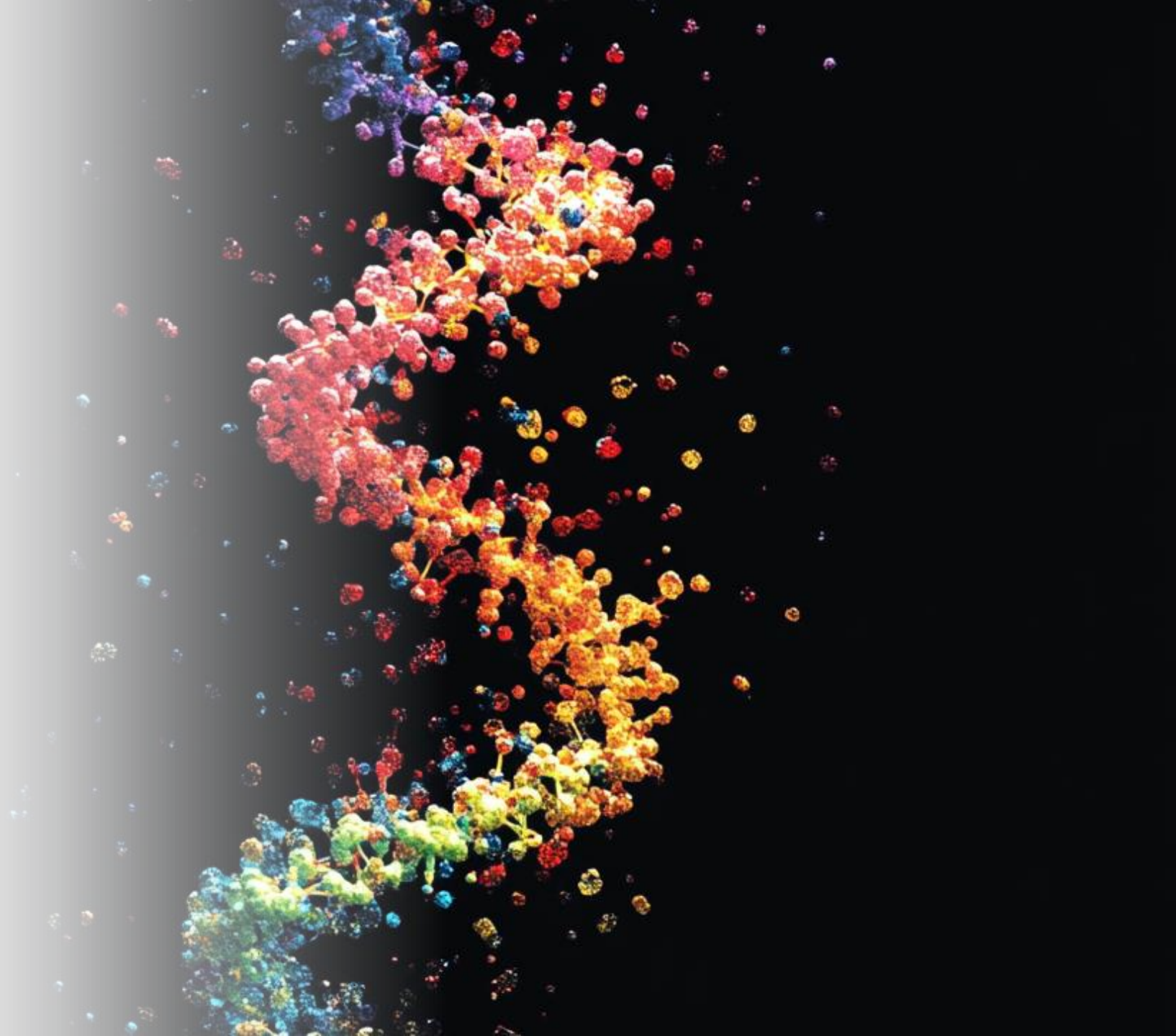
# RNA 3D Structure Prediction



# RNA 3D Structure Prediction

## Early Attempts

- First began in 1960
- “in silico”

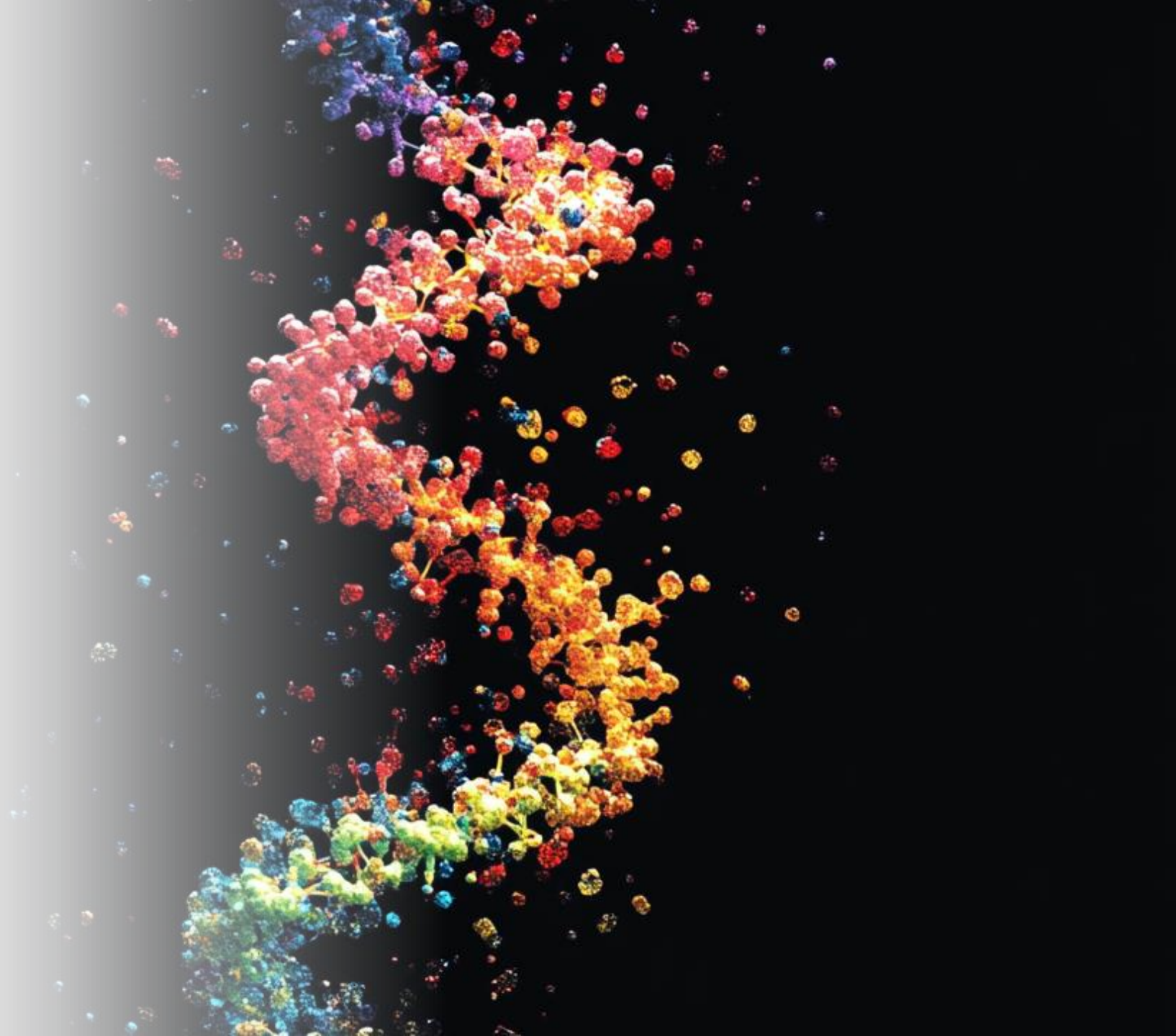




# RNA 3D Structure Prediction

## Early Attempts

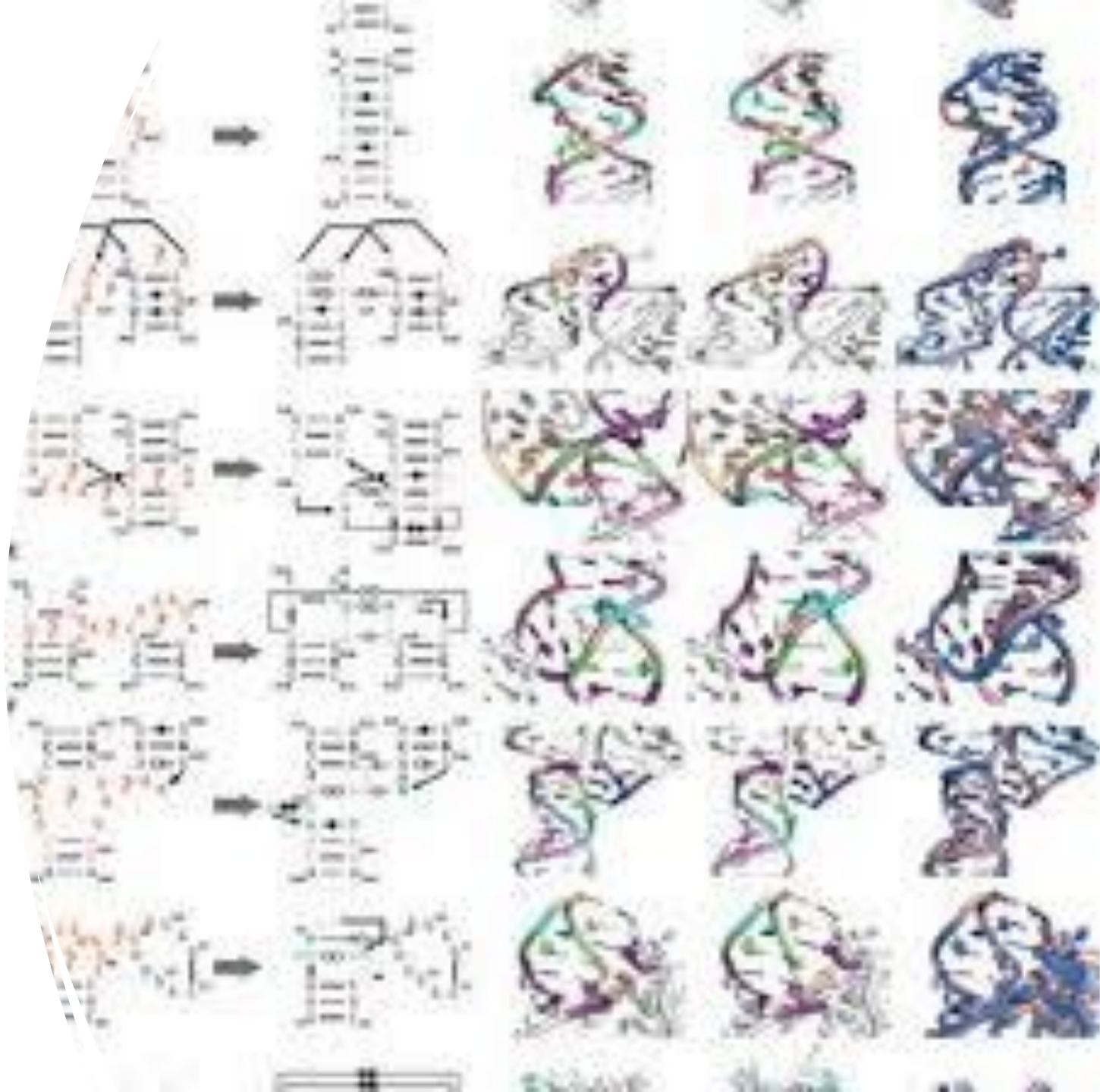
- First began in 1960
- “in silico”
- First interactive tool 1998



# Ab initio

---

- FARFAR
- iFoldRNA
- NAST
- SimRNA
- Vfold



# Homology Modeling

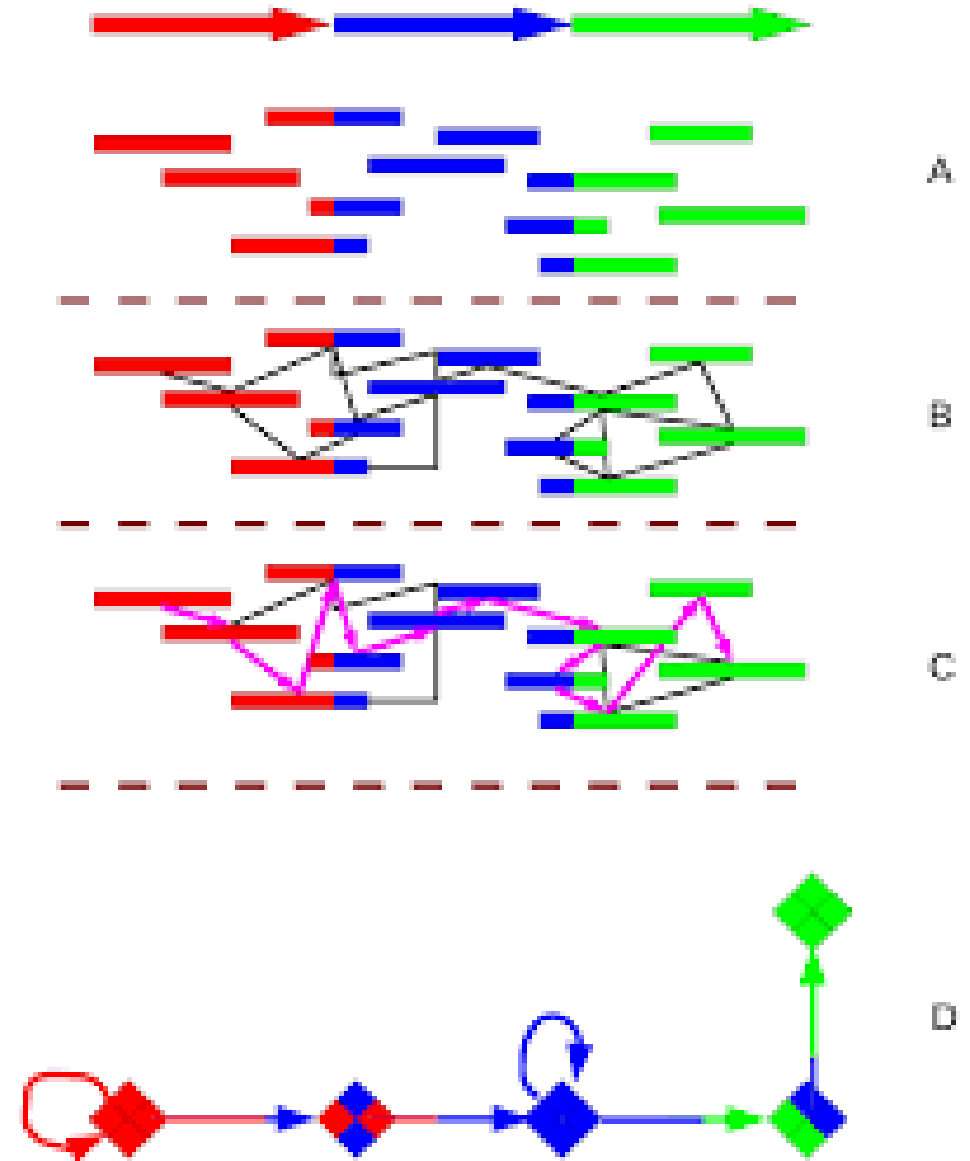
---

- RNABuilder
- ModelRNA



# Fragment-based Assembly

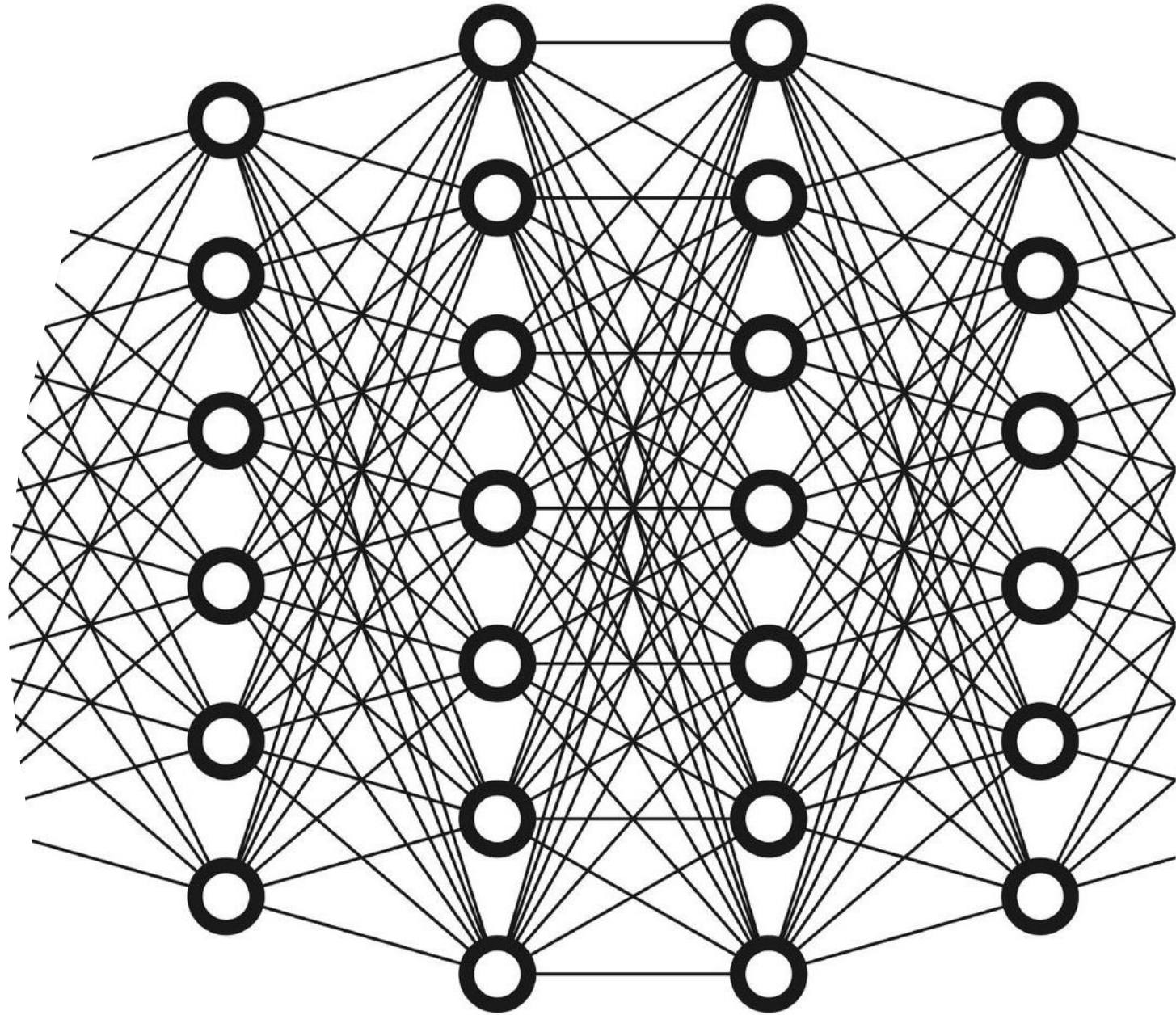
- MC-Fold/MC-Sym
- Assemble
- RNAComposer
- 3dRNA



# Deep Learning

---

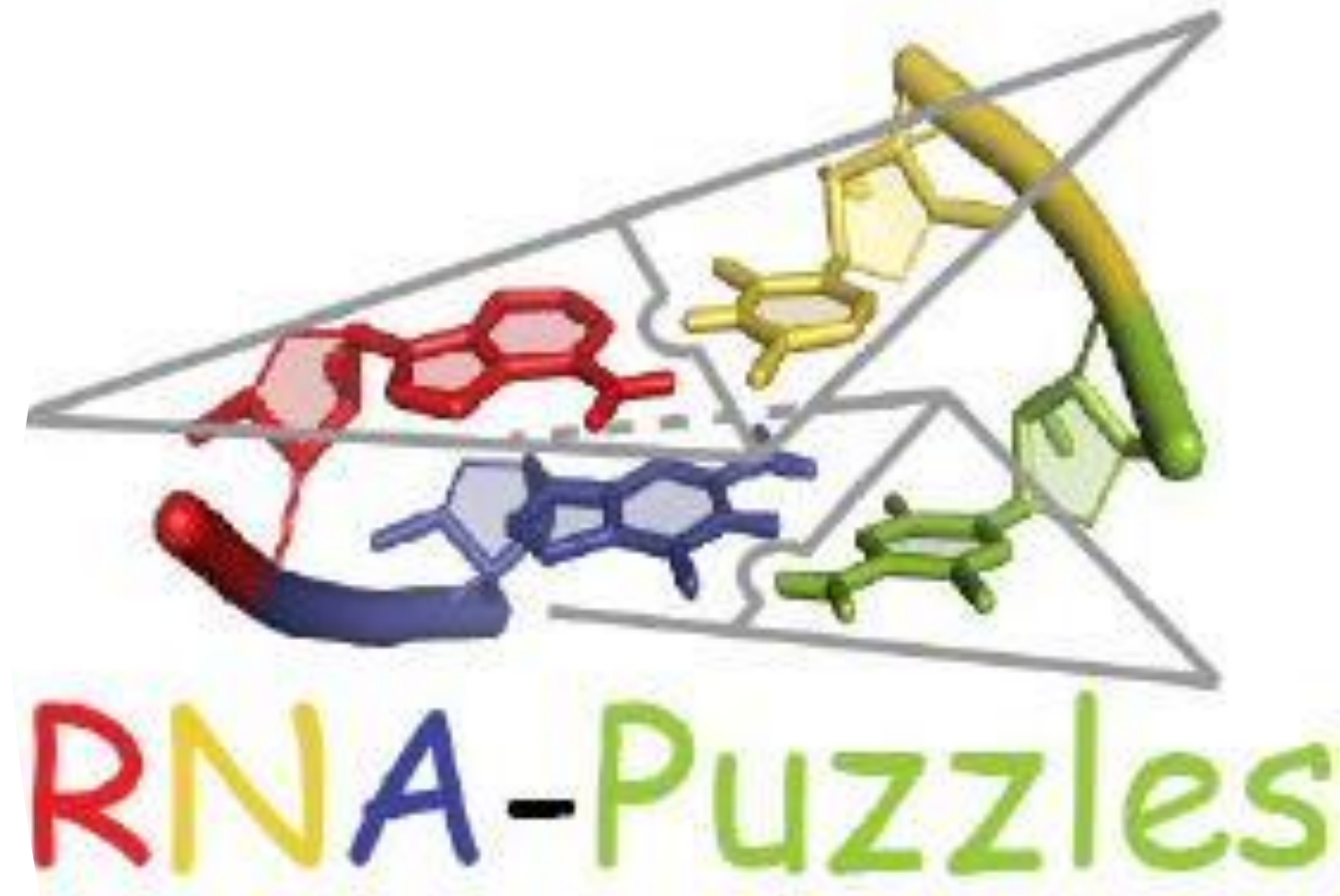
- Townshend et al.
- End-to-end prediction



# RNA-Puzzles: Evaluating RNA Structure Predictions

---

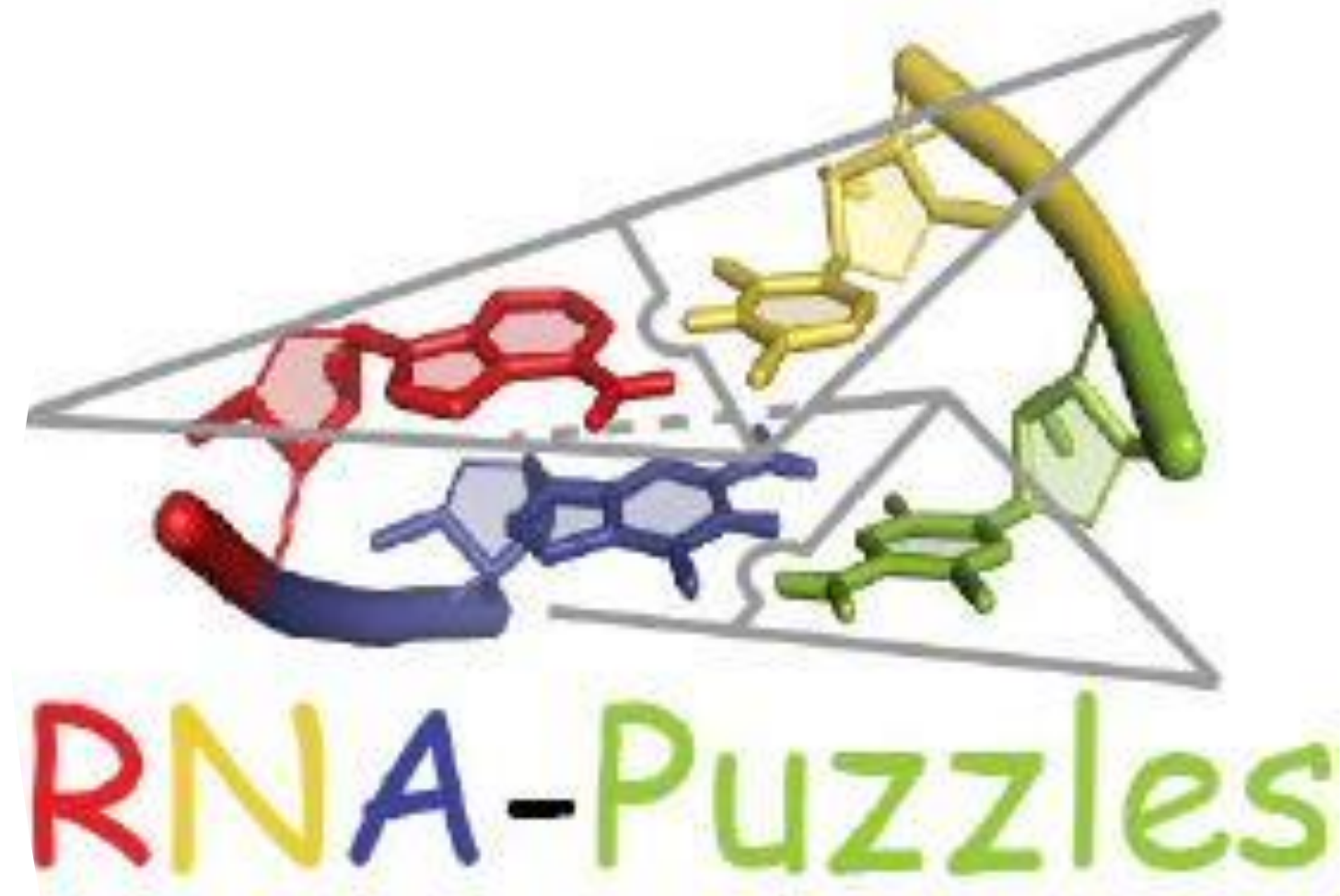
- Launched in 2010
- 38 challenges
- 2 projects



# RNA-Puzzles: Evaluating RNA Structure Predictions

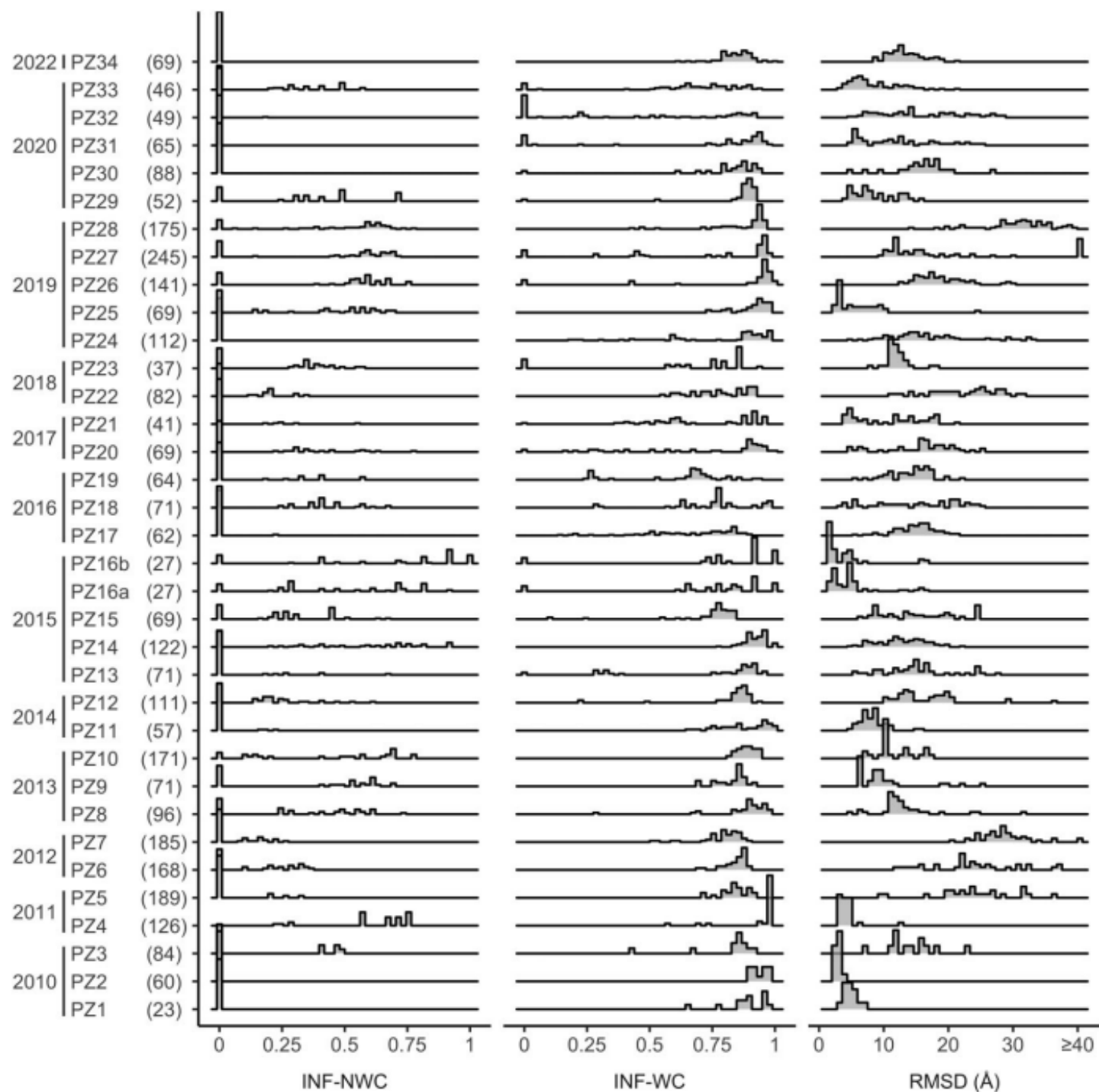
---

- Compared against reference structure
- Several similarity metrics
  - Interaction Network Fidelity (INF)
    - INF-WC
    - INF-NWC
    - INF-stacking



# 12 Years of Challenges

- INF-WC: 0.75 - 1.0
- INF-NWC:  $\sim 0$

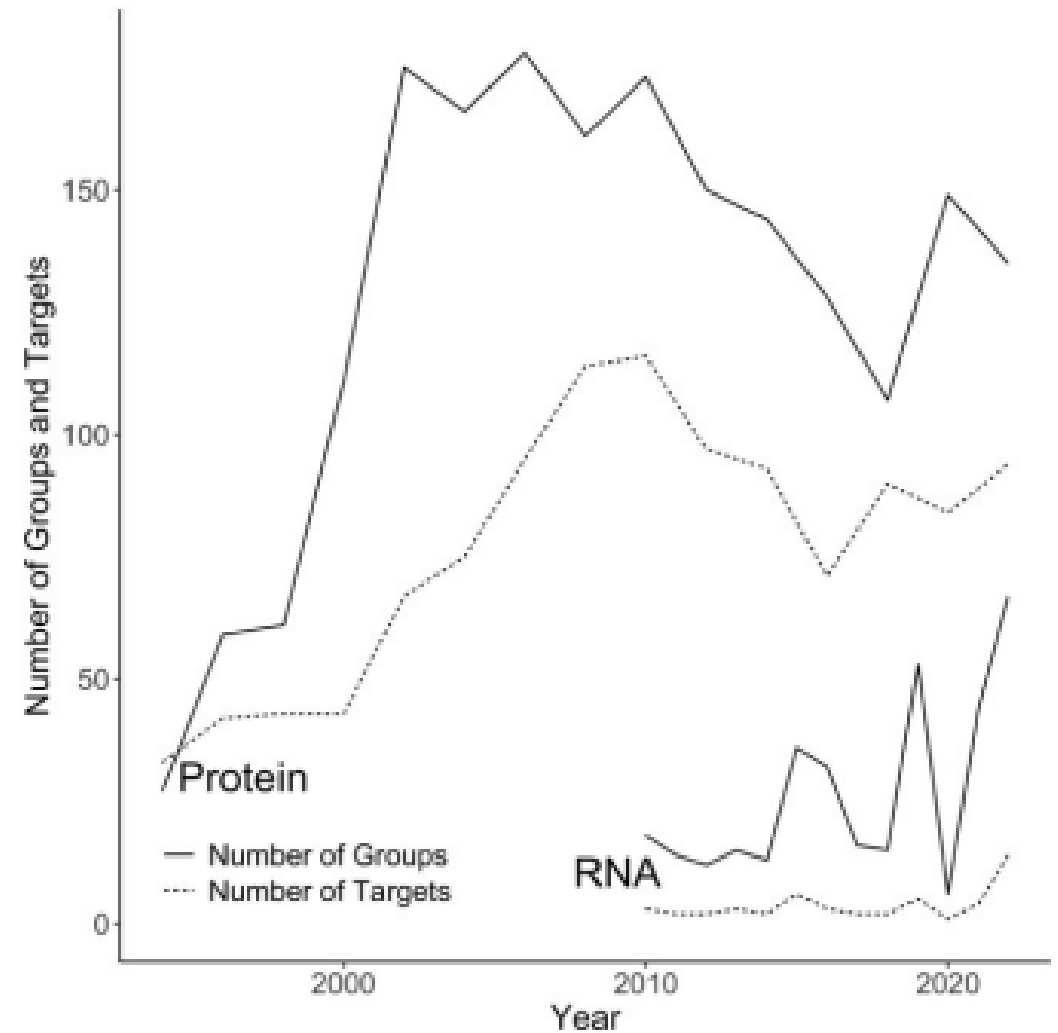




# CASP Recap

---

- CASP launched (1994)
- Fully automatic web-server prediction (2000)
- AlphaFold enters (2020)



**Figure 3.** Numbers of RNA and protein structure predictions made in RNA-Puzzles and CASP competitions. The solid lines represent the numbers of groups competing in CASP and RNA-Puzzles; the dashed lines are for the number of protein/RNA targets. From 2010 to 2021, RNAs were predicted only in RNA-Puzzles and in 2022, CASP included also RNA targets, which is responsible for the recent spike in targets and groups involved in 3D RNA structure prediction.



## CASP-RNA

- RNA Puzzles web server (2015)
- Deep learning models (2022)
- CASP-RNA (2022)
  - 42 groups
  - 12 RNA sequences
  - 18 teams used DL models
    - DeepFoldRNA
    - RhoFold
    - trRosettaRNA
    - OpenComplexRNA
- **Top 4 were non-ML**

---

# The Challenges



# Low Quality and Quantity

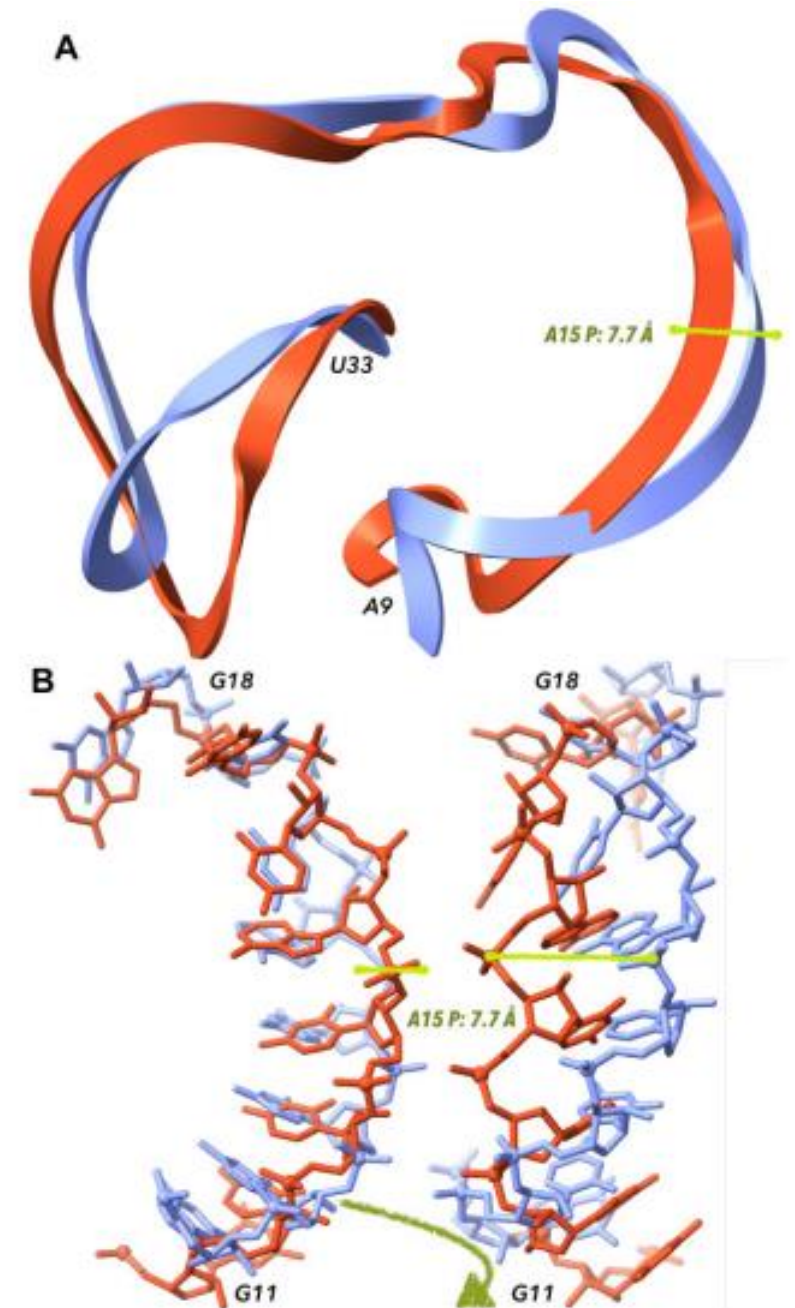
**Table 1.** Numbers of all PDB-released structures (\*) and residues in X-ray and cryo-EM structures (\*\*) with high resolution ( $\leq 2.0 \text{ \AA}$ ) over decades. In the first column, amino acids are abbreviated as AAs, and nucleotides as nts

	$\leq 1980$	1981–1990	1991–2000	2001–2010	2011–2022	Total	% of the total
Proteins (*)	78	634	12 121	43 205	108 677	164 715	91.57
AAs $\leq 2.0 \text{ \AA}$ (**)	5050	45 236	1 609 401	11 390 238	28 513 777	41 563 702	99.78
RNA (*)	2	23	306	1392	4488	6211	3.45
RNA nts $\leq 2.0 \text{ \AA}$ (**)	0	0	1270	5974	26 921	34 165	0.08
DNA (*)	1	91	1061	2009	5800	8962	4.98
DNA nts $\leq 2.0 \text{ \AA}$	0	238	5430	15 730	38 107	59 505	0.14

# Complexity of RNA Architecture

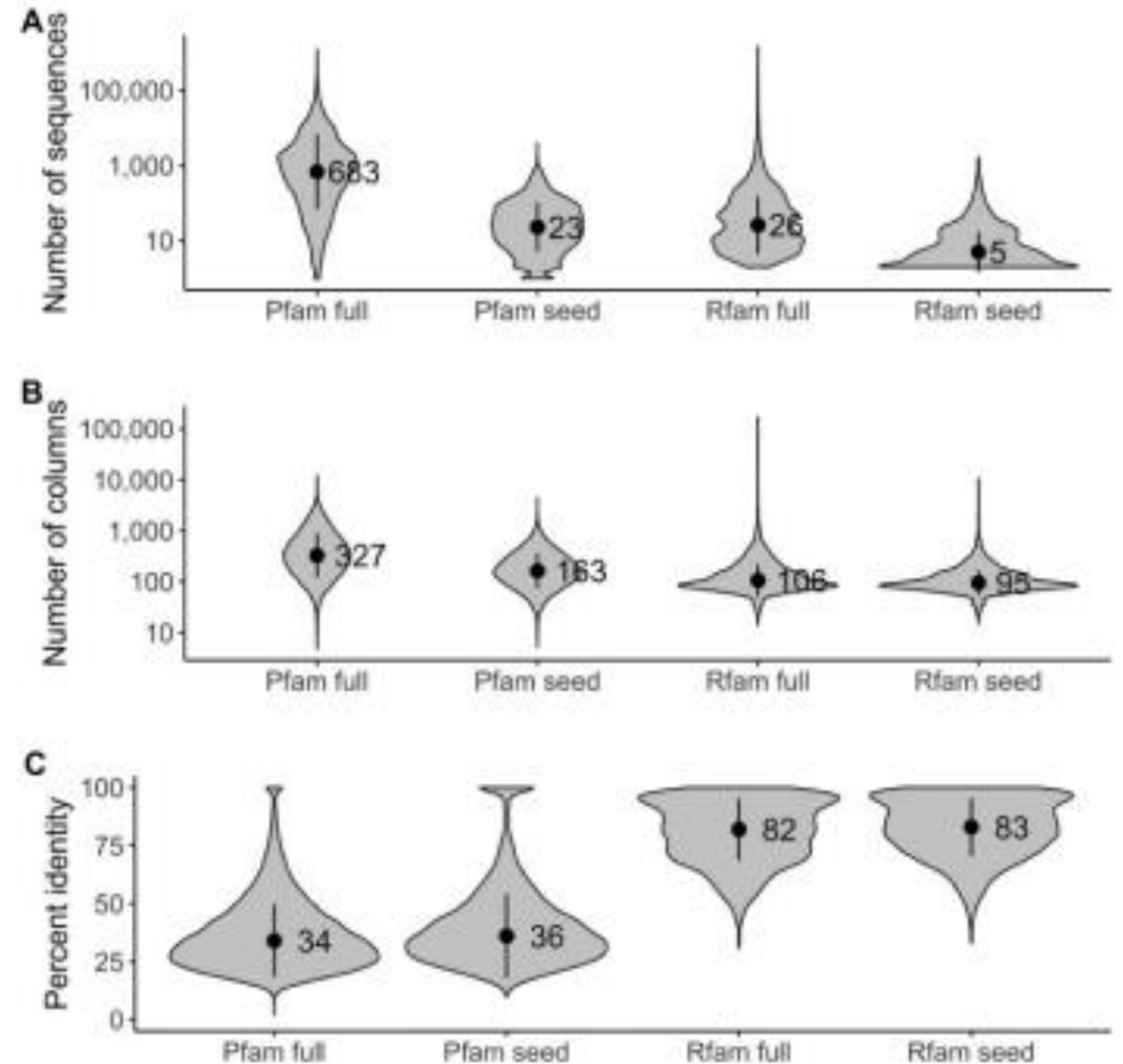
---

- A-form antiparallel double helix is the easiest
- RNA junctions are complex and challenging
- Scarce data for non-Watson-Crick base pairs
- Inter-/Intra-molecular interactions



# Limited Sequence and Alignment Data

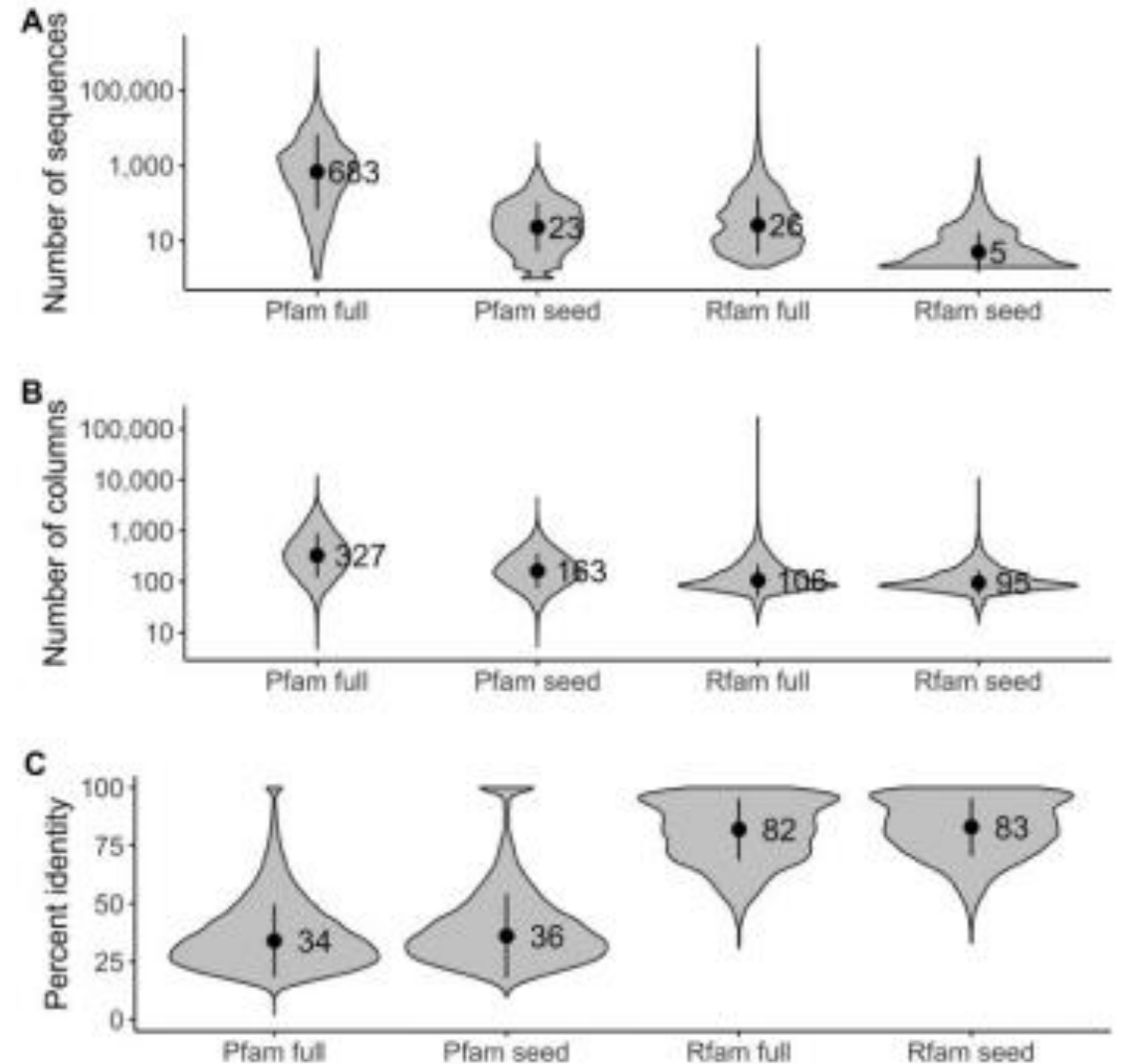
---



**Figure 5.** Rfam versus Pfam alignments compared based on (A) a number of sequences, (B) a number of columns and (C) the average pairwise percent identity for each family. The points on the plots indicate the mean, and the vertical bars indicate the standard deviation.

# Challenges with use of RNA Alignments

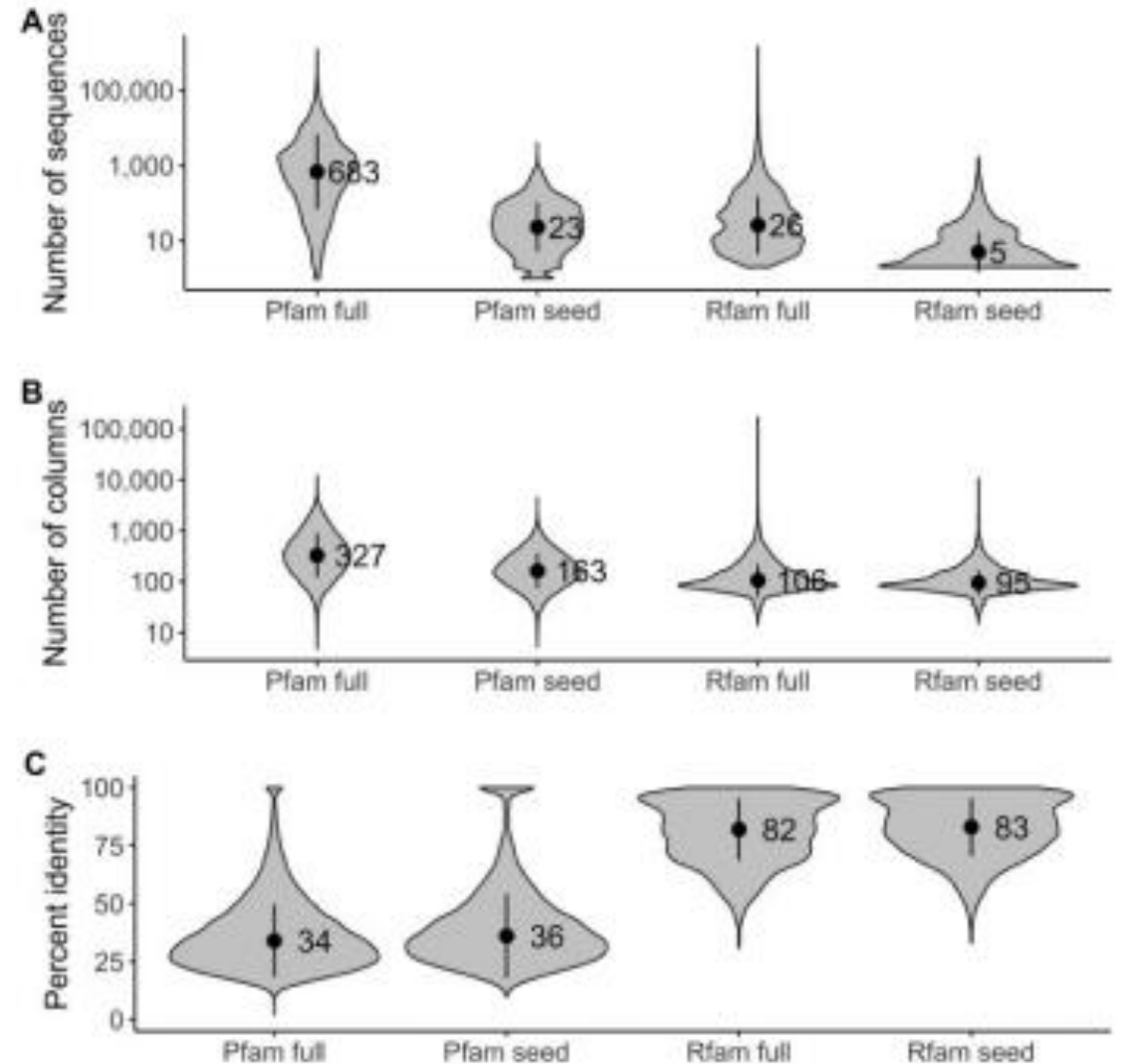
1) Rfam construction is expensive



**Figure 5.** Rfam versus Pfam alignments compared based on (A) a number of sequences, (B) a number of columns and (C) the average pairwise percent identity for each family. The points on the plots indicate the mean, and the vertical bars indicate the standard deviation.

# Challenges with use of RNA Alignments

- 1) Rfam construction is expensive
- 2) RNA alignments are smaller



**Figure 5.** Rfam versus Pfam alignments compared based on (A) a number of sequences, (B) a number of columns and (C) the average pairwise percent identity for each family. The points on the plots indicate the mean, and the vertical bars indicate the standard deviation.



# Challenges with use of RNA Alignments

- 1) Rfam construction is expensive
- 2) RNA alignments are smaller
- 3) Rfam has global biases

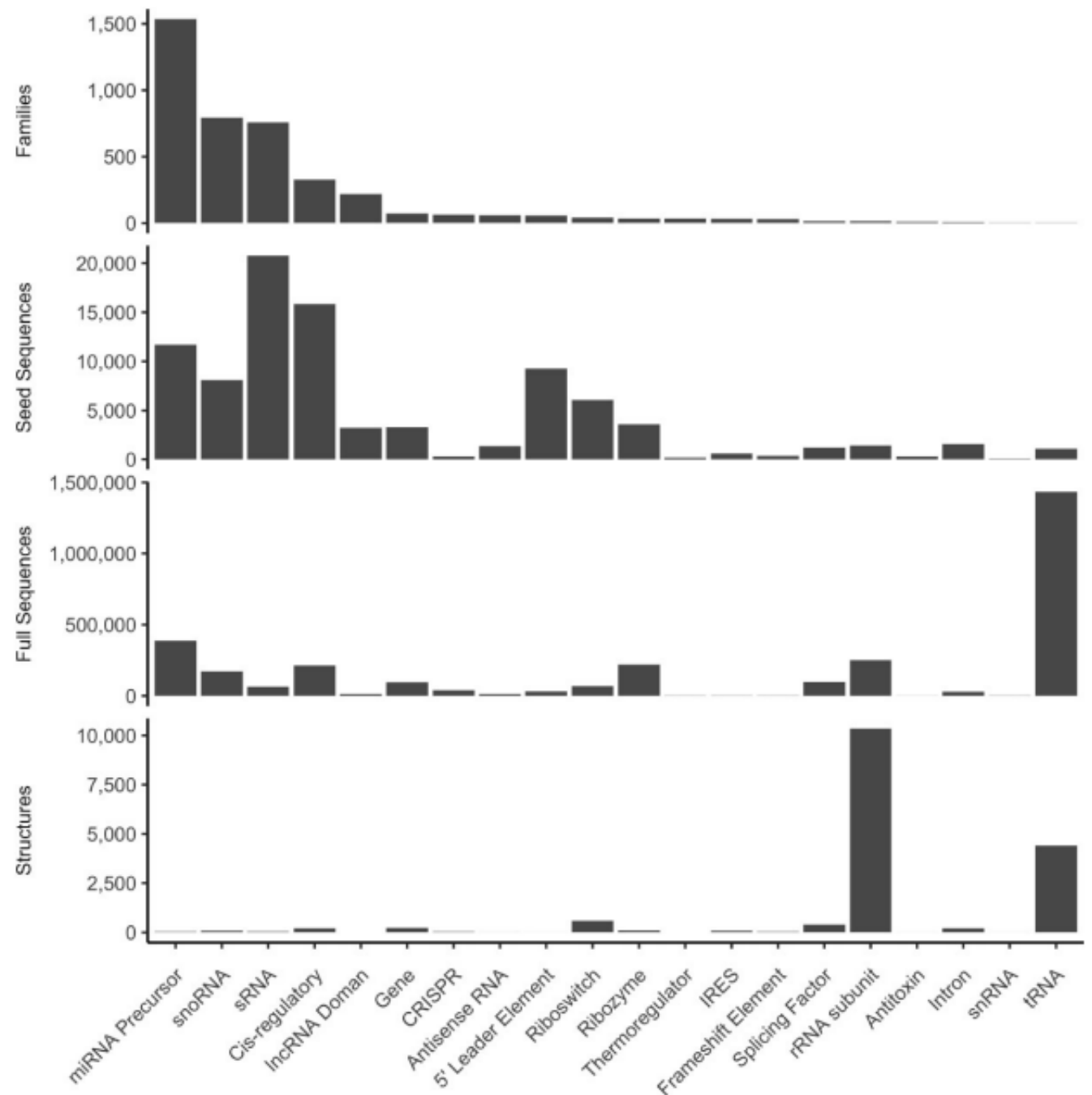


Figure 6. Counts of Rfam families, seed sequences, full sequences and structures for all Rfam families organized by Rfam RNA type.

# Challenges with use of RNA Alignments

---

- 1) Rfam construction is expensive
- 2) RNA alignments are smaller
- 3) Rfam has global biases
- 4) Inconsistency with NWC base pairs



# Challenges with use of RNA Alignments

---

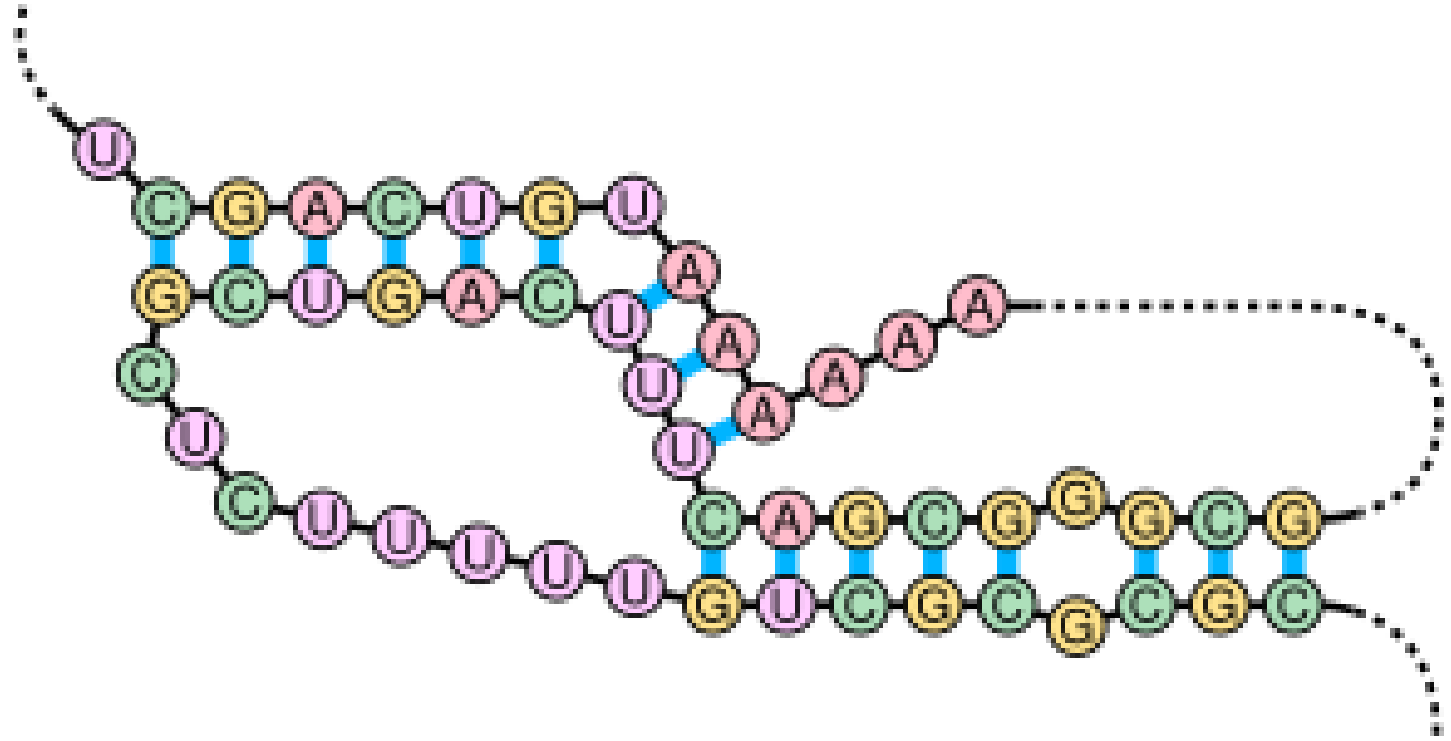
- 1) Rfam construction is expensive
- 2) RNA alignments are smaller
- 3) Rfam has global biases
- 4) Inconsistency with NWC base pairs
- 5) Secondary structure representation



# Challenges with use of RNA Alignments

---

- 1) Rfam construction is expensive
- 2) RNA alignments are smaller
- 3) Rfam has global biases
- 4) Inconsistency with NWC base pairs
- 5) Secondary structure representation
- 6) Lack of pseudoknots





# Challenges with use of RNA Alignments

- 1) Rfam construction is expensive
- 2) RNA alignments are smaller
- 3) Rfam has global biases
- 4) Inconsistency with NWC base pairs
- 5) Secondary structure representation
- 6) Lack of pseudoknots



# The Future RNA Prediction Landscape



# Speculations

- AlphaFold took 50 years
- AlphaFold trained on 170,000 proteins
- Pfam has 19,632 alignments
- Rfam will have 19,000 in 70 years



# Improvements

- RNA motifs and global structures
- More research on molecular interactions
- Improved benchmarking
- Improved MSA
- Consistent annotation of RNA families
- Incorporation of new technologies





Final Call to Action



Thank You!

Questions?

The background of the image is a dense, overlapping collage of numerous small, rectangular pieces of paper. Each piece of paper is a light, muted color (ranging from pale blue to light beige) and features a large, dark, stylized question mark printed on it. The papers are scattered across the entire frame, creating a textured, layered effect. A thin, vertical white line runs down the center of the image, starting from the top edge and extending to the bottom edge, passing behind the text.