

Geometric Deep Learning of RNA Structure

Townshend, Raphael JL, et al. Science 373.6558 (2021): 1047-1051.

Hyeonseo Hwang, Advanced Bioinformatics



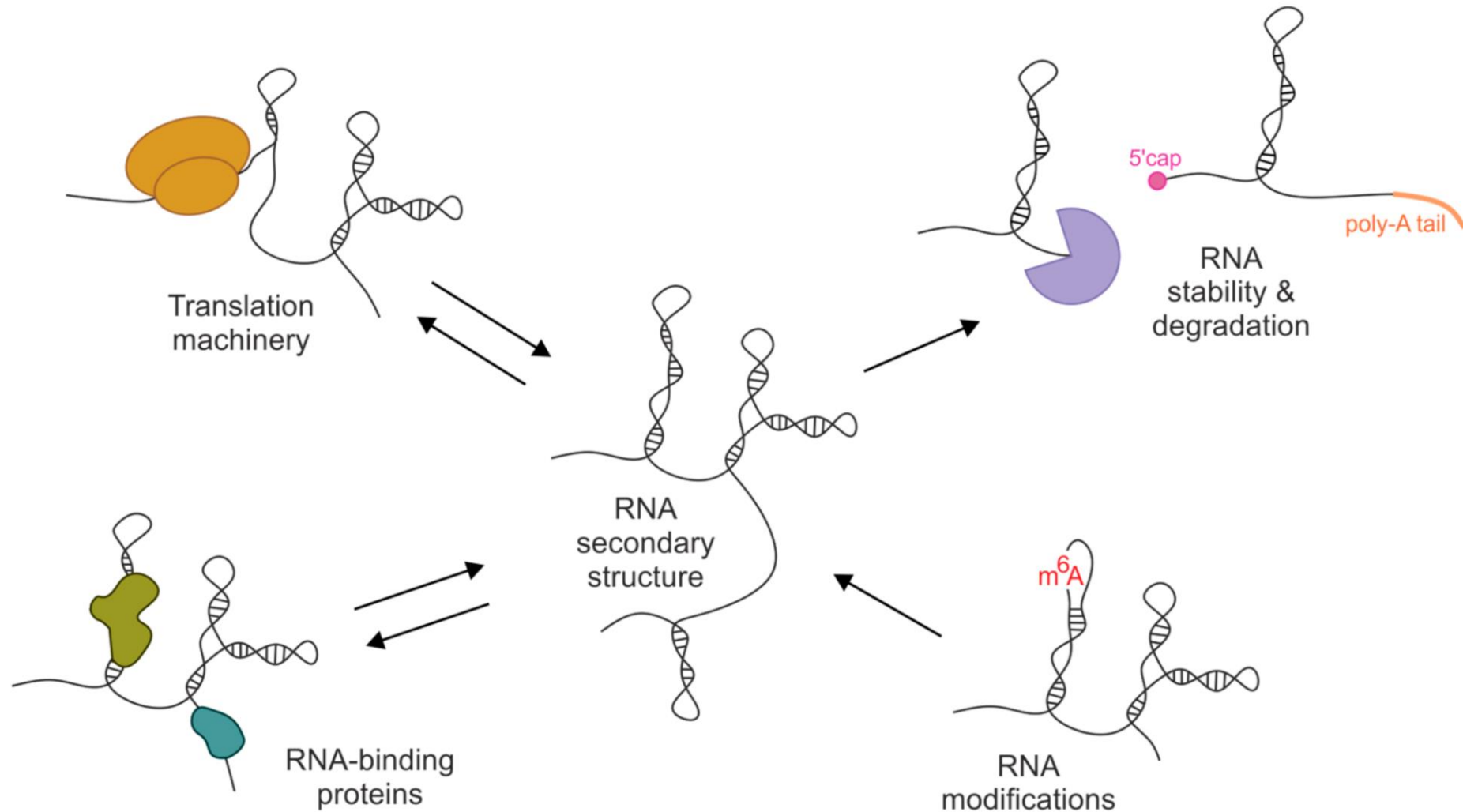
```
graph LR; A[Background] --> B[Architecture & Training]; B --> C[Benchmark]
```

Background

Architecture
& Training

Benchmark

1. RNA 3-D Structure has Biological Importance



(Andrzejewska, A. et al., 2020)

1. Two Strategies for RNA 3-D Structure Prediction

Comparative
Sequence Analysis

Folding Algorithms
+
Scoring Schemes

Experi-
mental

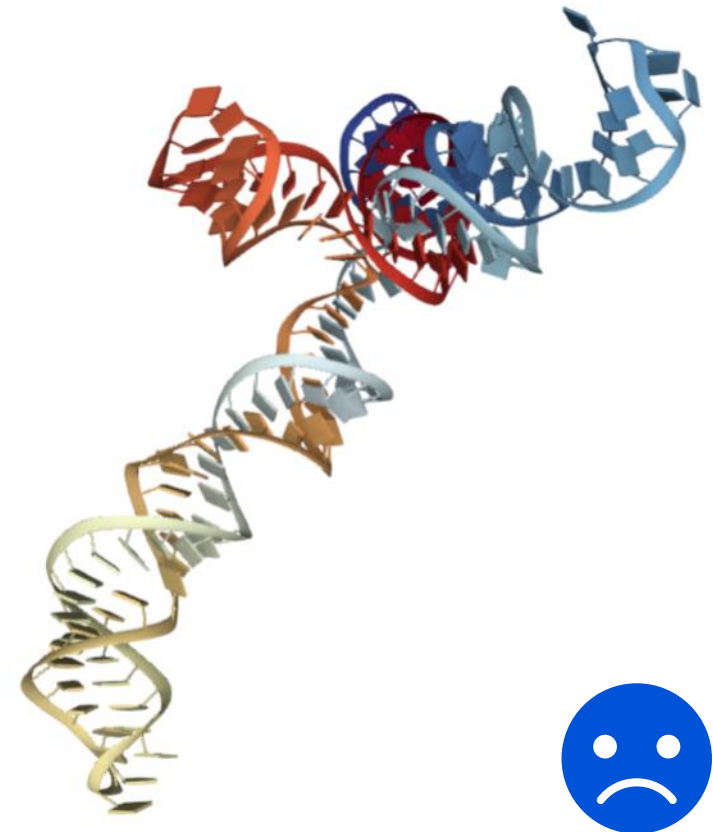
Machine
Learning

1. RNA 3-D Structure Prediction is Challenging

Robust Prediction
(AlphaFold2)

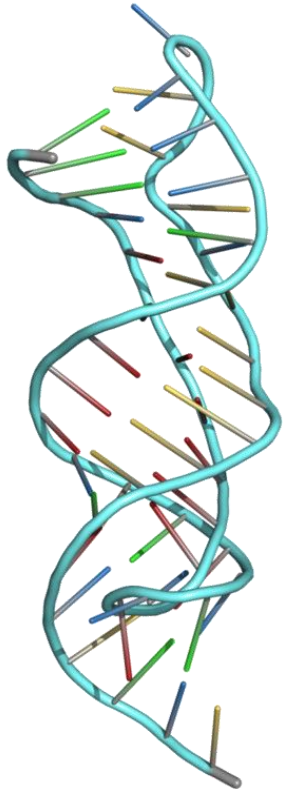


Performance ceiling at
~80% Precision

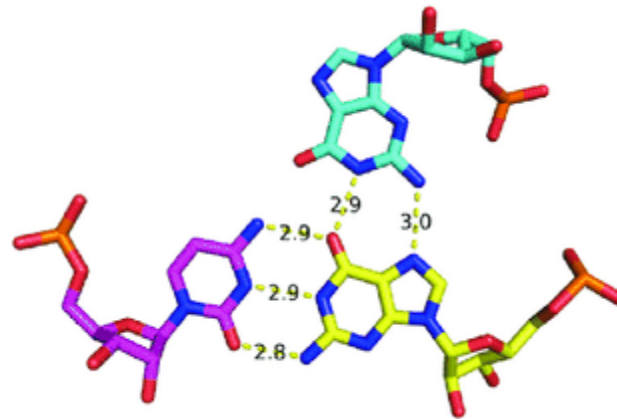


1. Why is it Difficult to Predict RNA Structures?

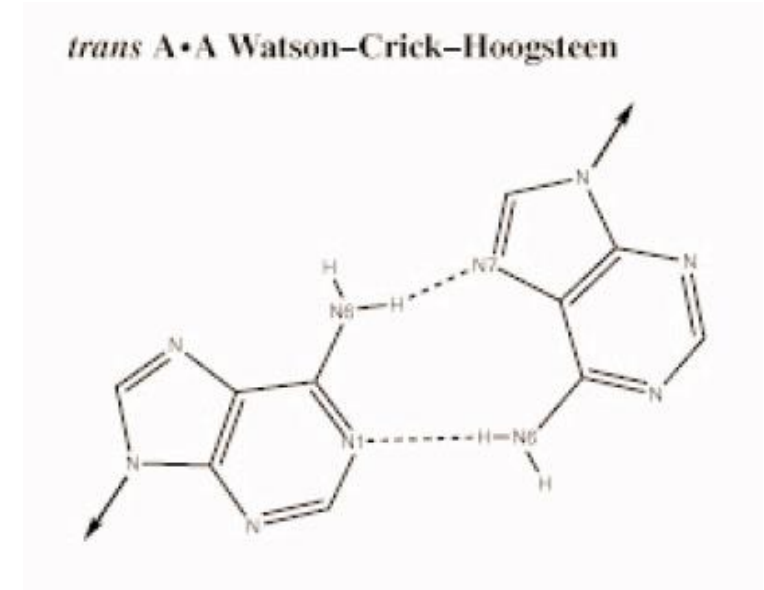
Results of tertiary interaction



Pseudoknot



Triplet Interaction

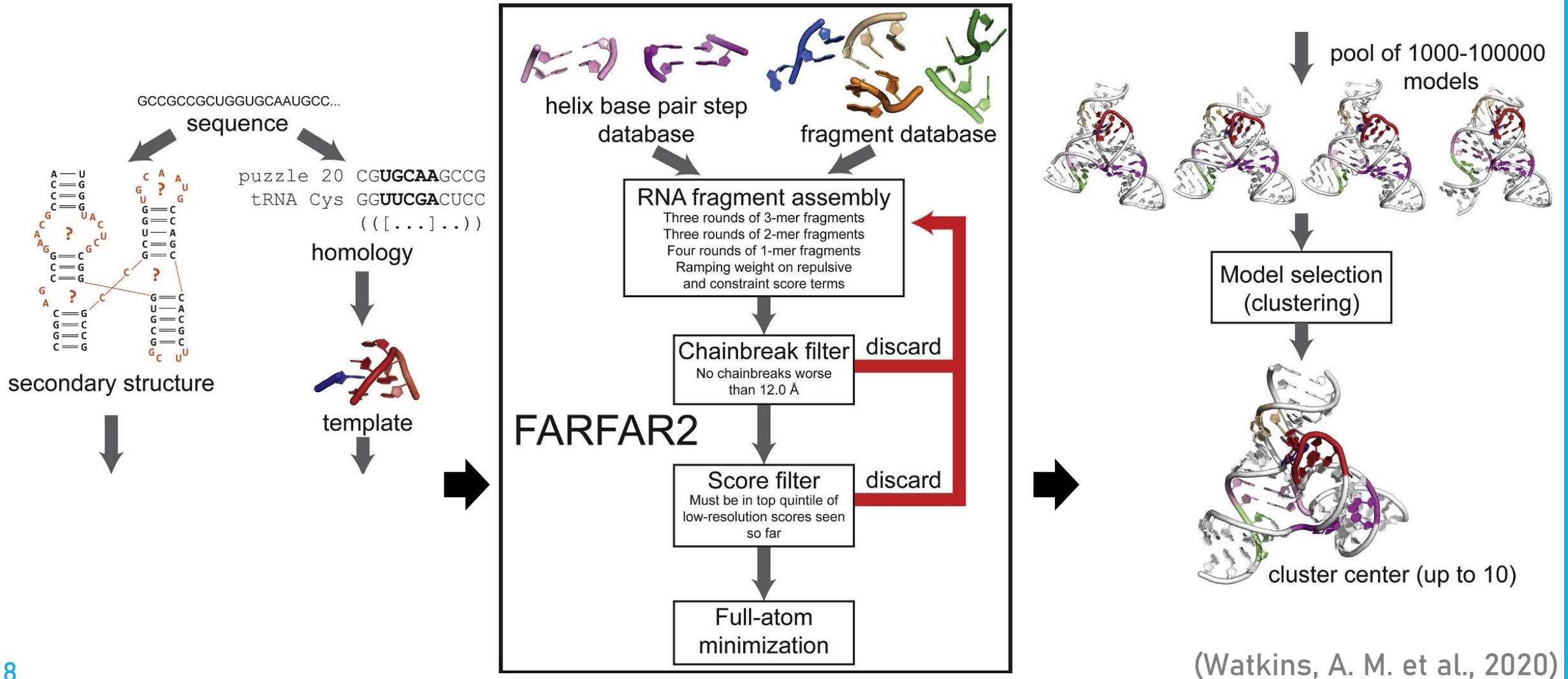


Noncanonical Pairing

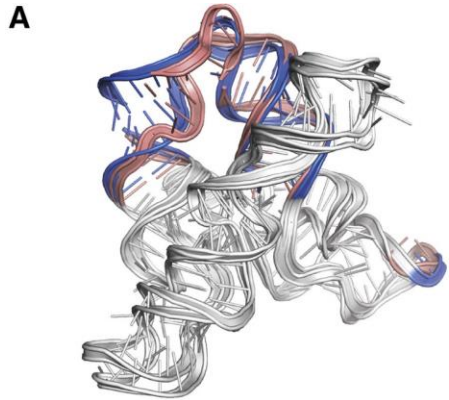
1. Why is it Difficult to Predict RNA Structures?

- Low availability of closely related template structures.
- Low value of sequence coevolution information.
- Characteristics of energetically favorable RNA structures are not well understood.

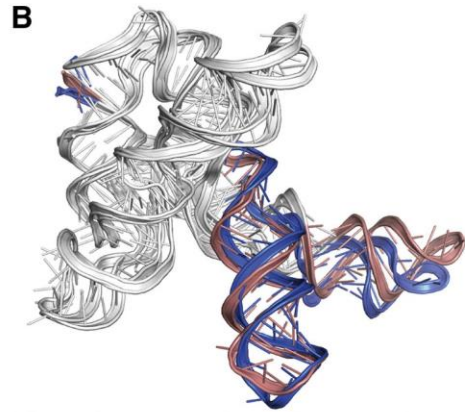
2. FARFAR2 Predicts RNA 3-D Structure



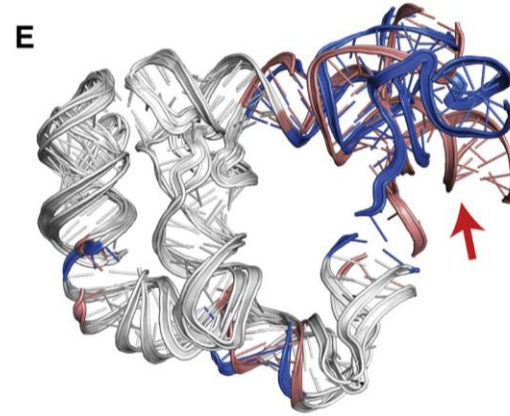
2. FARFAR2 Predicts RNA 3-D Structure



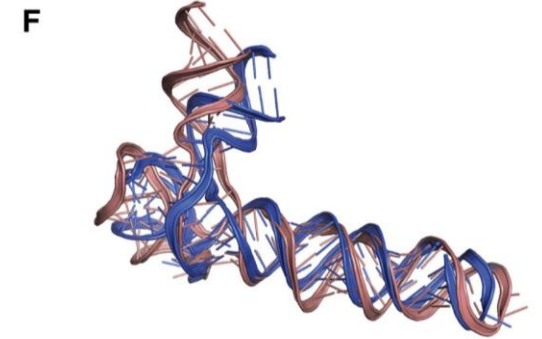
F. nucleatum tandem glycine riboswitch



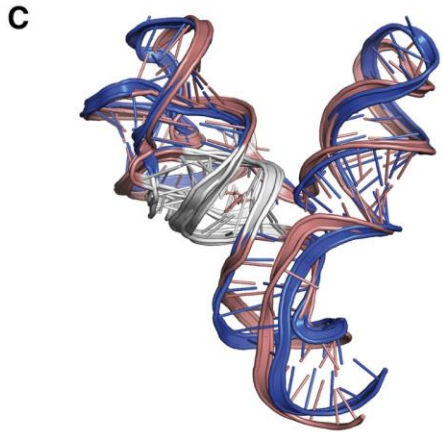
V. cholerae tandem glycine riboswitch



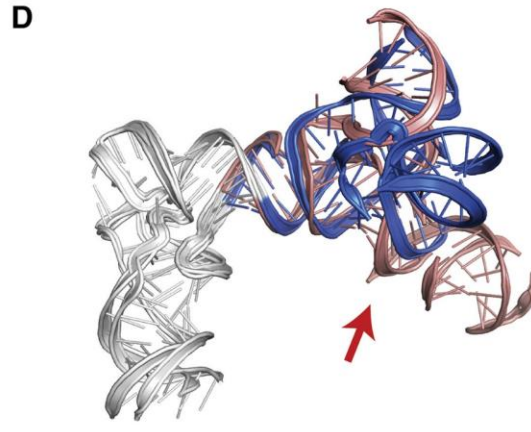
B. subtilis T-box riboswitch



VA RNA I



Mycobacterium SAM-IV riboswitch



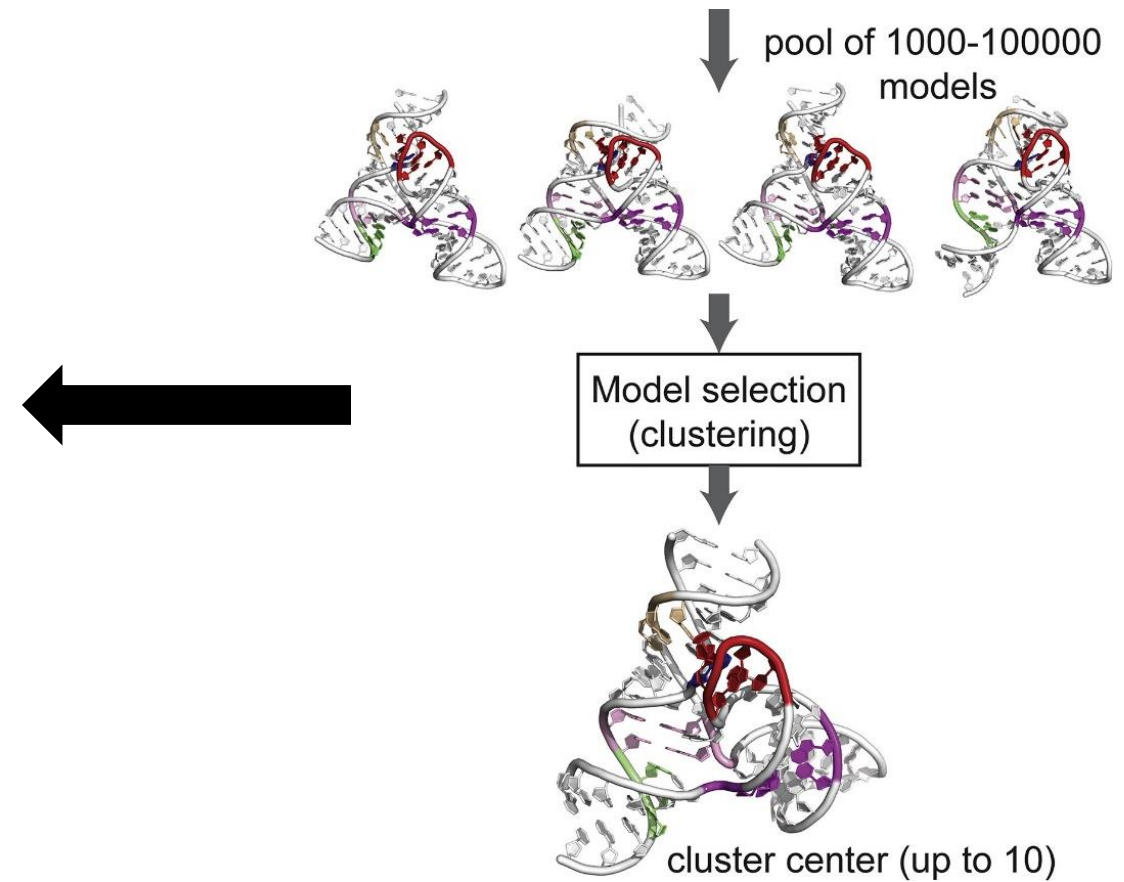
G. kaustophilus T-box riboswitch

There are rooms for improvement.

(Watkins, A. M. et al., 2020)

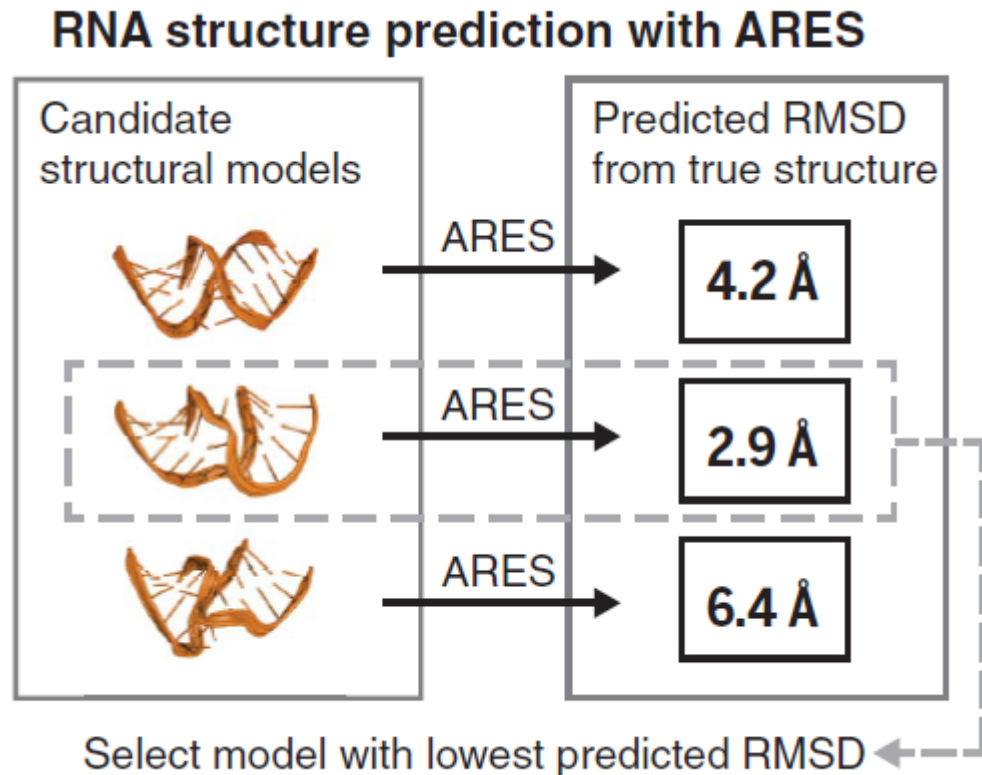
2. FARFAR2 Predicts RNA 3-D Structure

How can we find the **BEST** model from the pool of thousand structures?



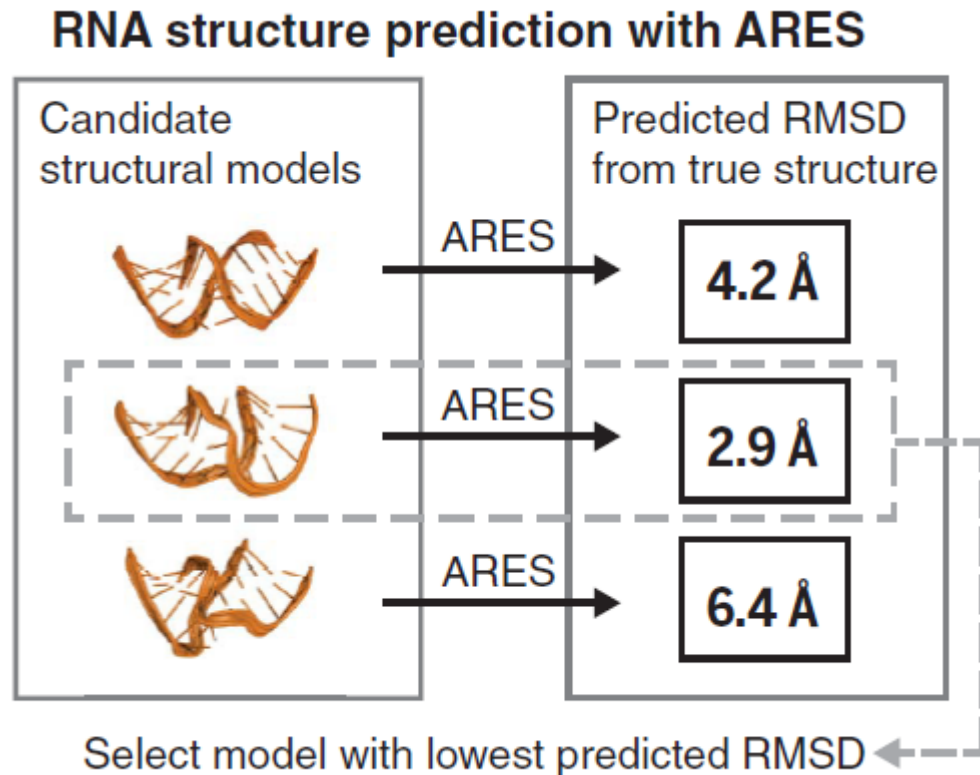
(Watkins, A. M. et al., 2020)

3. ARES Scores Candidate RNA Structures



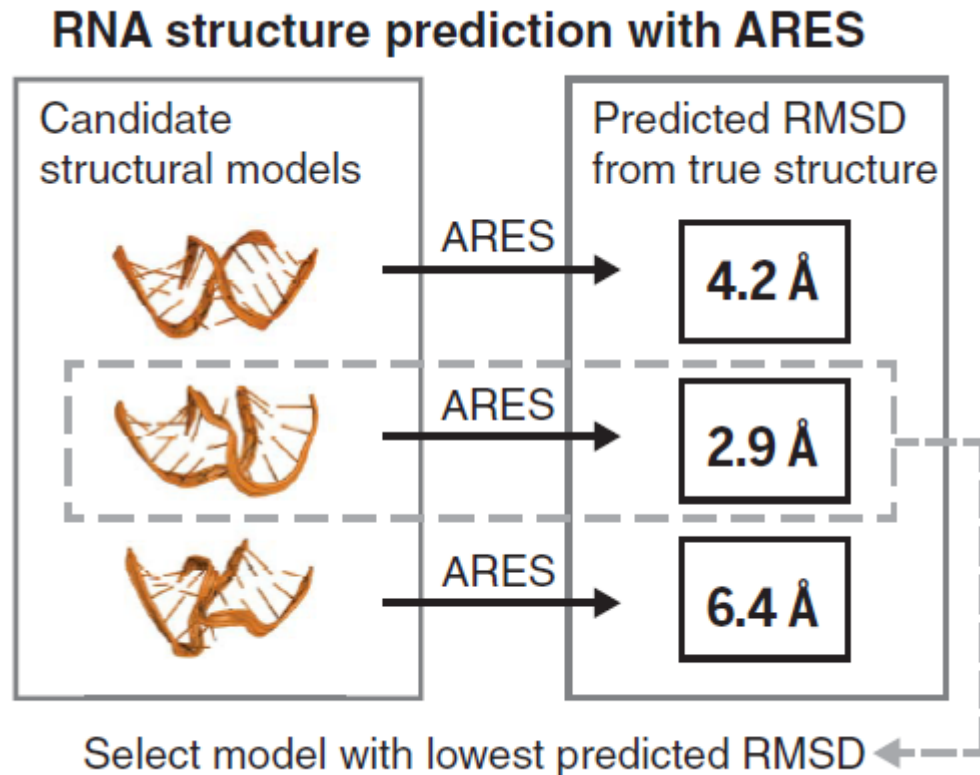
Atomic
Rotationally
Equivariant
Scorer

3. ARES Scores Candidate RNA Structures



Learns from known RNA structures to assess the accuracy of RNA structures.

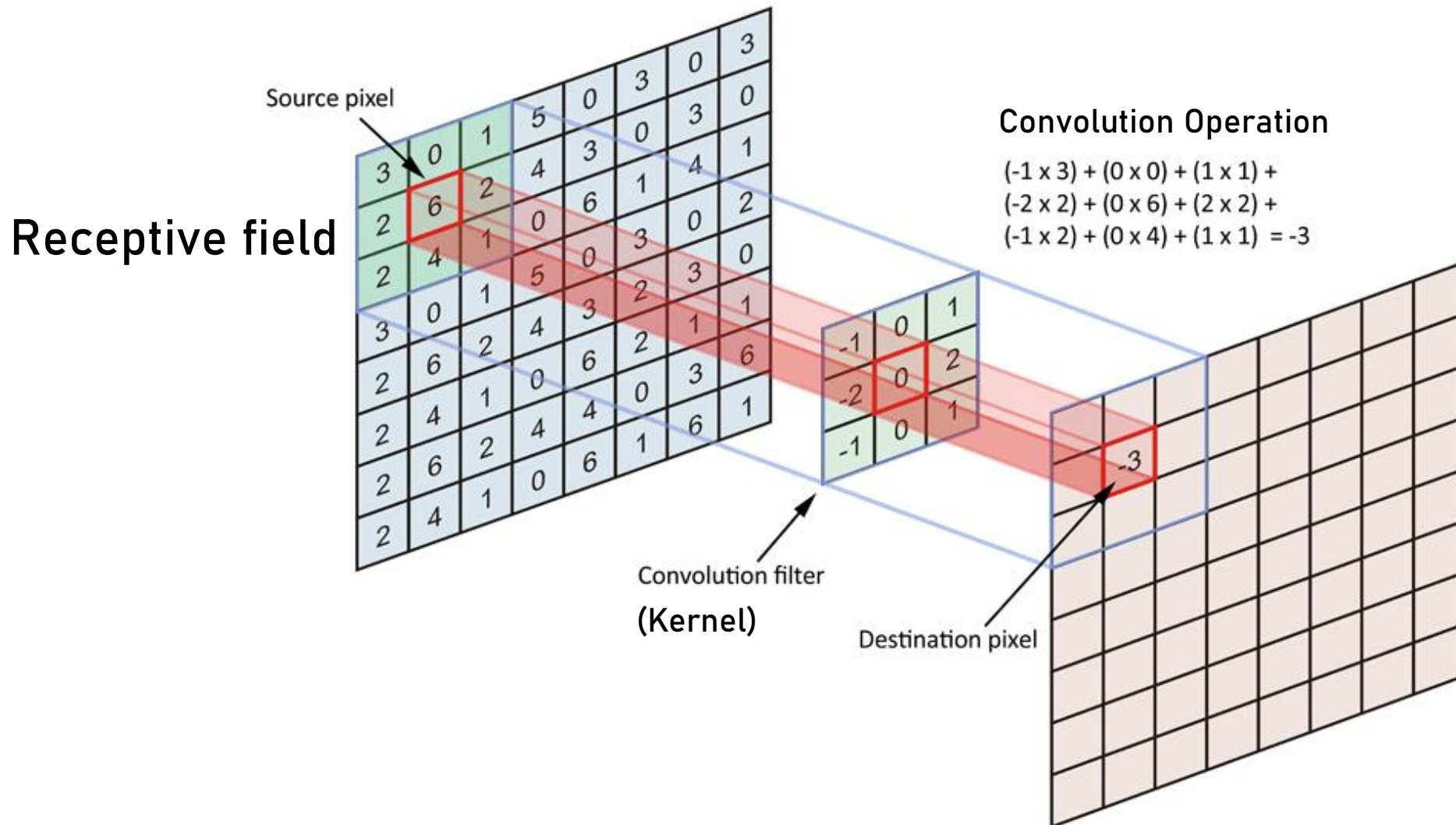
3. Challenges in Scoring Candidate RNA Structures



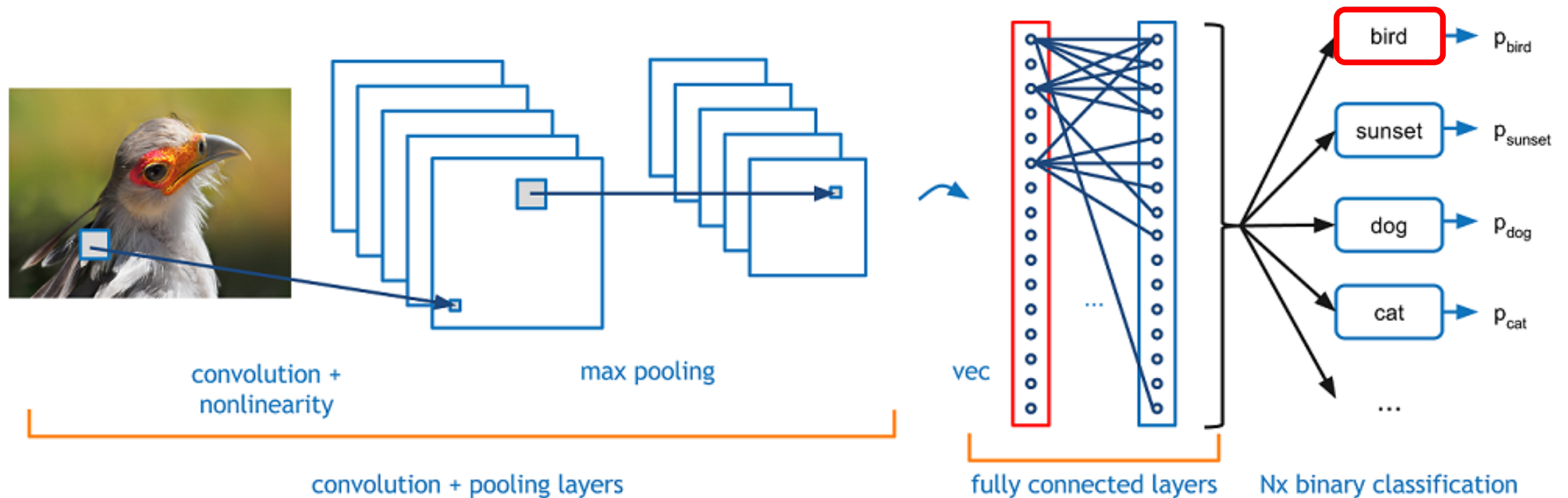
ARES has to:

- Avoid assumptions about known structural characteristics.
- Learn from the limited number of RNA structures.

4. Convolutional Neural Network (CNN)

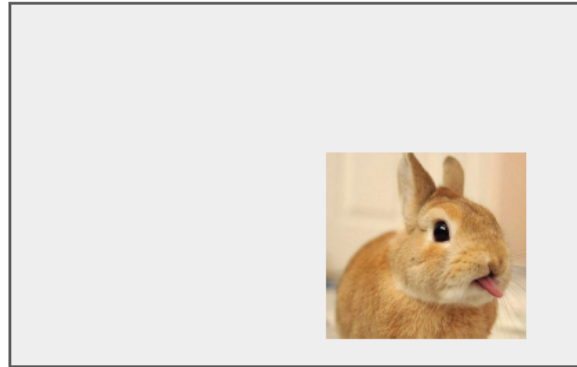
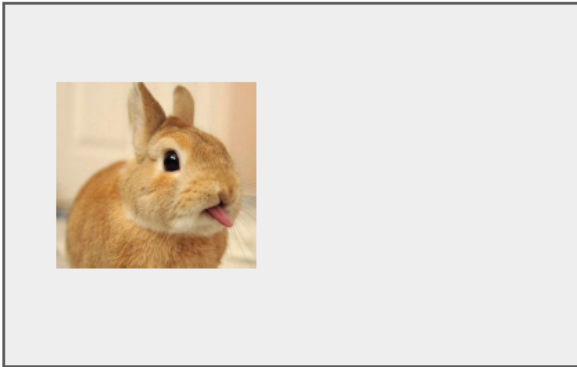


4. Convolutional Neural Network (CNN)



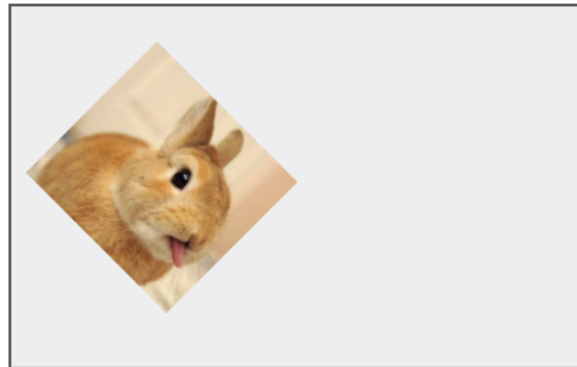
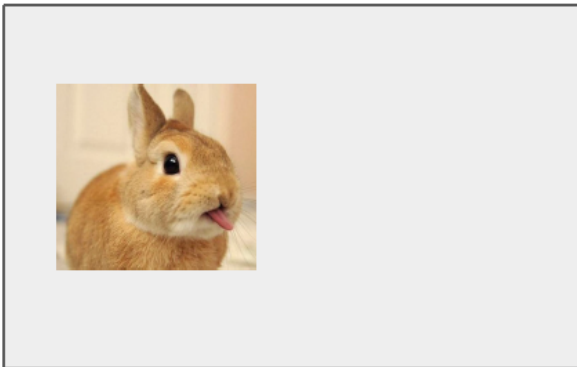
Each neuron gets input from its receptive field.

4. Tensor Field Network and Rotation Equivariance



Translation equivariance

Convolutional neural network ✓

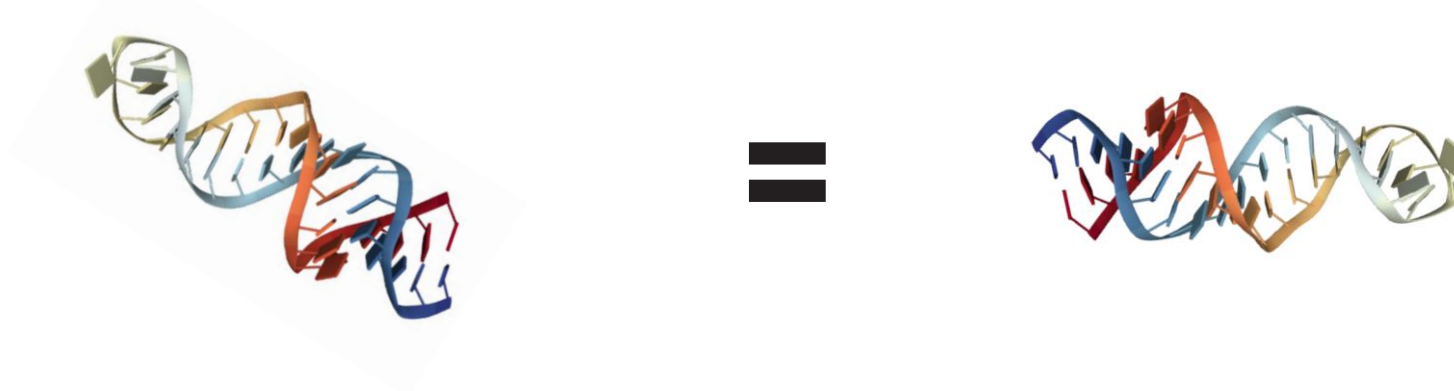


Rotation equivariance?

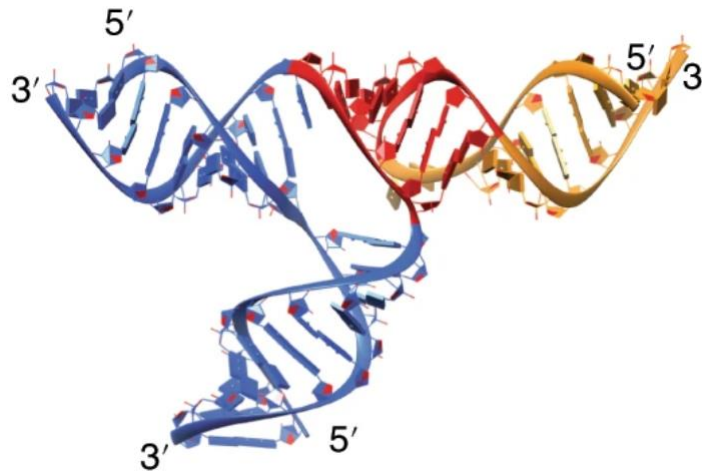
(Smidt, T., 2018)

4. Why is Rotation Equivariance Important?

Invariance of physics to rotation or translation



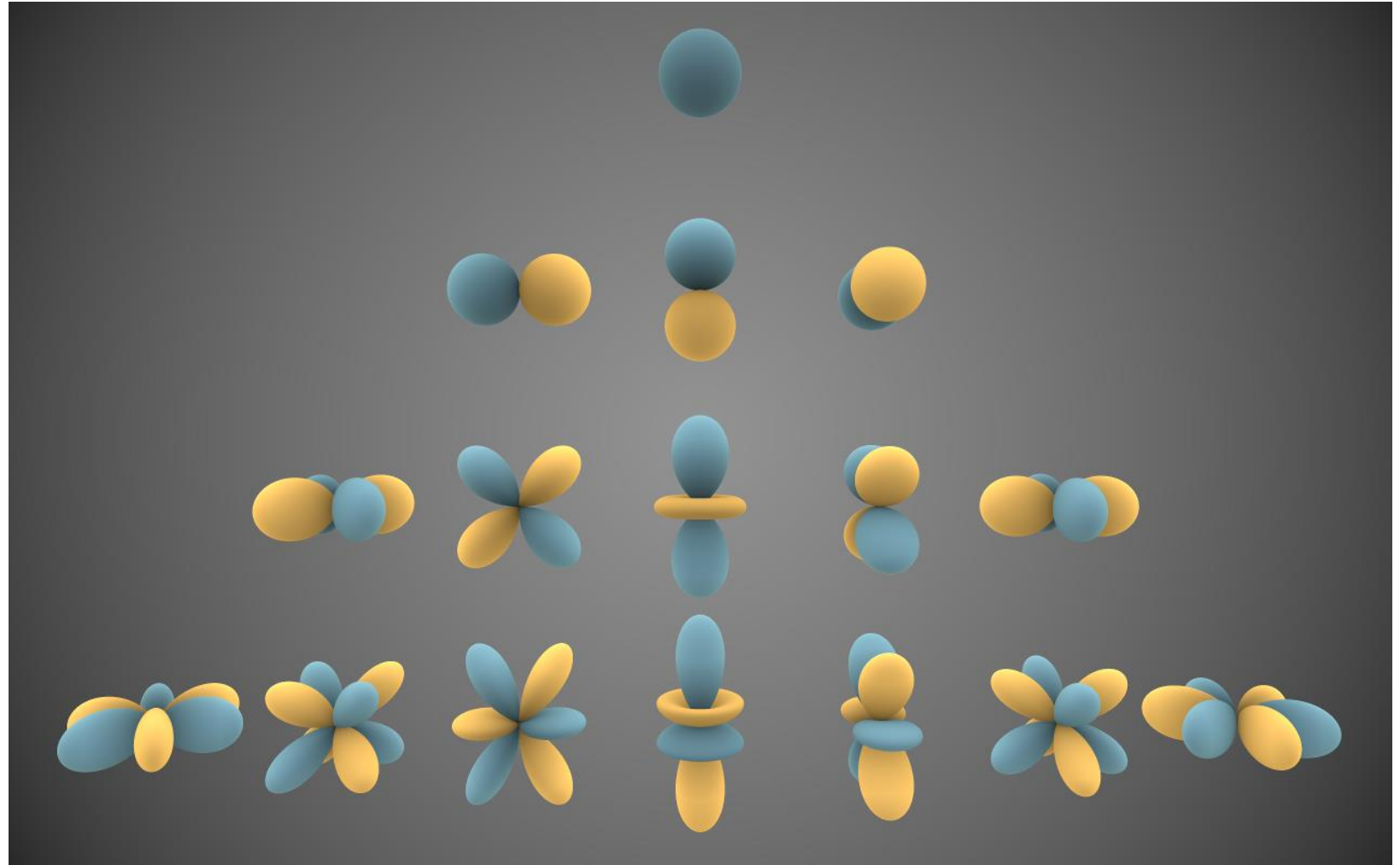
Preserve orientation and position → Recognise larger motifs.



4. Tensor Field Network and Rotation Equivariance

Spherical Harmonics

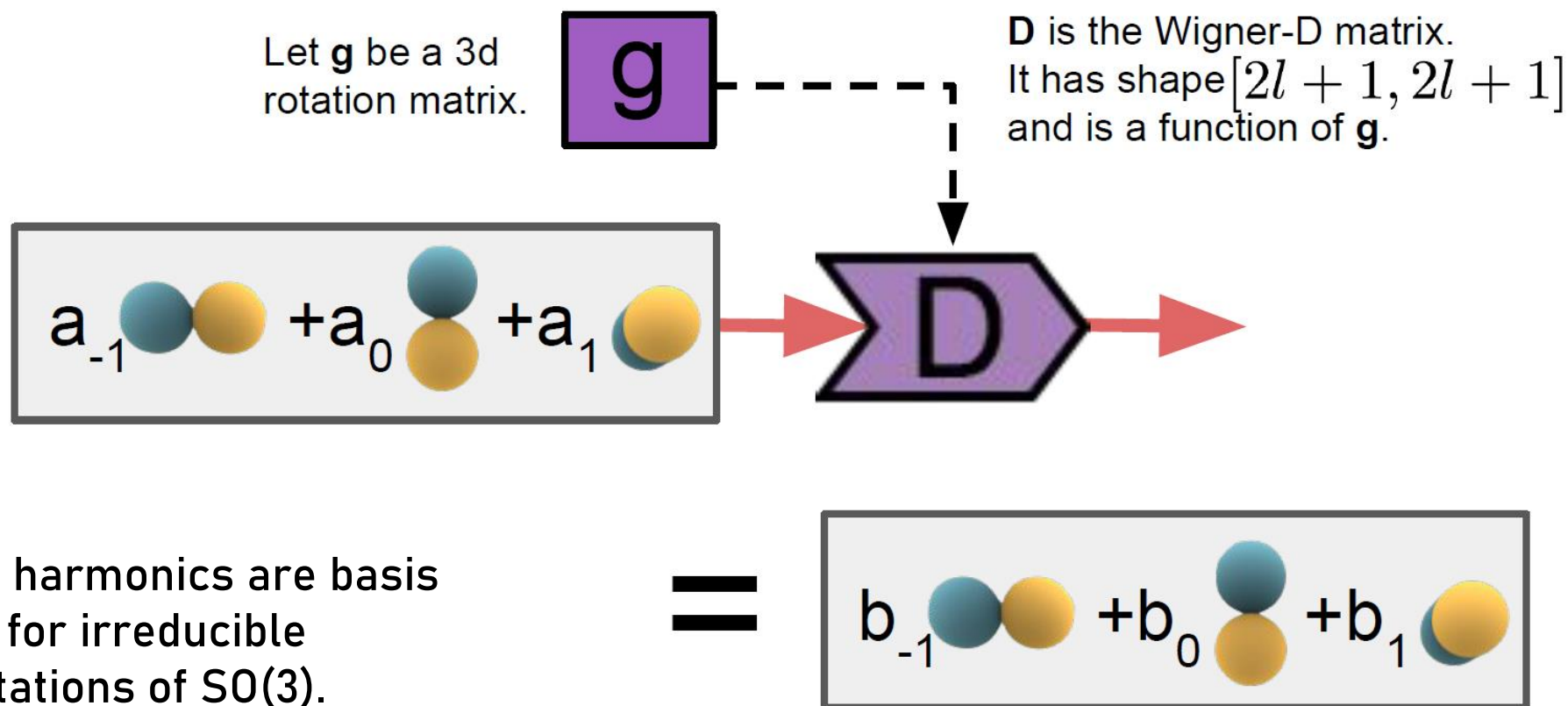
- Eigenfunctions of the angular part of the Laplacian in 3-D



(Wikimedia)

4. Tensor Field Network and Rotation Equivariance

Spherical harmonics of a given L transform together under rotation.



Spherical harmonics are basis functions for irreducible representations of $SO(3)$.

(Smidt, T., 2018)

Background



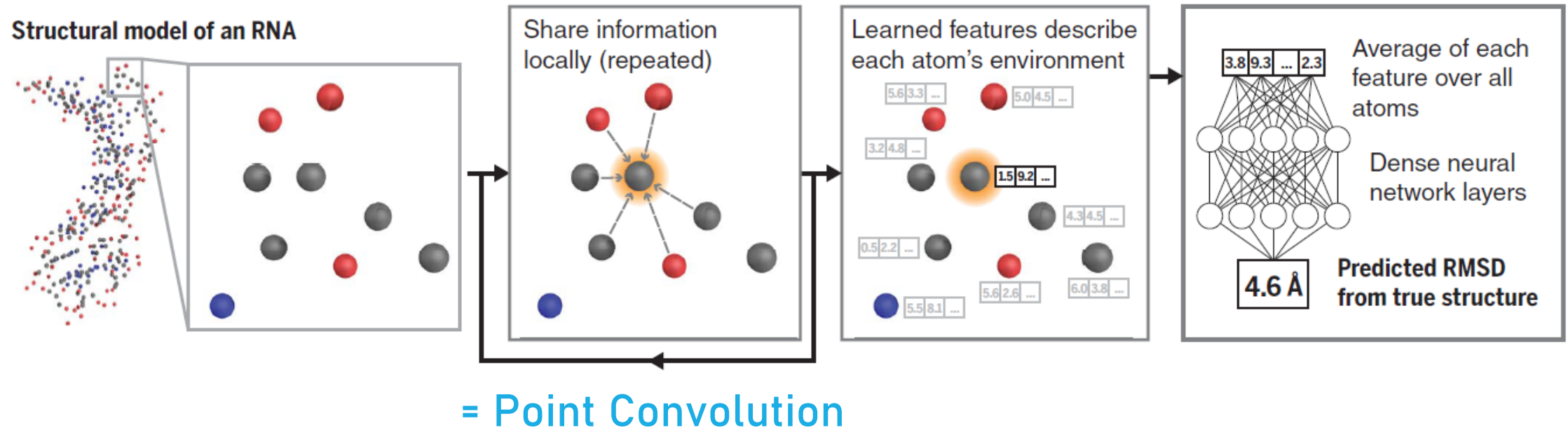
**Architecture
& Training**



Benchmark

5. ARES Architecture: Overview

A ARES predicts the accuracy of a structural model, given only atomic coordinates and element types

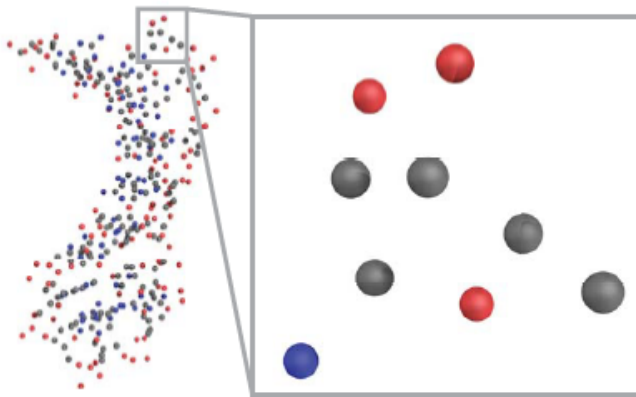


5. ARES Only Receives Atomic Information

Structural Model Input File (.pdb)

A ARES predicts the accuracy of a st

Structural model of an RNA



| | | | | | | | | | | | | |
|----|------|----|------|---|---|----|---------|--------|--------|------|------|---|
| 1 | ATOM | 1 | C5' | G | A | 13 | -18.149 | -2.986 | 72.212 | 1.00 | 0.00 | C |
| 2 | ATOM | 2 | C4' | G | A | 13 | -18.733 | -2.791 | 70.831 | 1.00 | 0.00 | C |
| 3 | ATOM | 3 | O4' | G | A | 13 | -20.018 | -2.119 | 70.941 | 1.00 | 0.00 | O |
| 4 | ATOM | 4 | C3' | G | A | 13 | -17.944 | -1.903 | 69.884 | 1.00 | 0.00 | C |
| 5 | ATOM | 5 | O3' | G | A | 13 | -16.865 | -2.590 | 69.259 | 1.00 | 0.00 | O |
| 6 | ATOM | 6 | C1' | G | A | 13 | -20.201 | -1.246 | 69.842 | 1.00 | 0.00 | C |
| 7 | ATOM | 7 | C2' | G | A | 13 | -19.019 | -1.468 | 68.899 | 1.00 | 0.00 | C |
| 8 | ATOM | 8 | O2' | G | A | 13 | -19.363 | -2.476 | 67.969 | 1.00 | 0.00 | O |
| 9 | ATOM | 9 | N1 | G | A | 13 | -20.904 | 3.631 | 68.639 | 1.00 | 0.00 | N |
| 10 | ATOM | 10 | C2 | G | A | 13 | -20.974 | 2.516 | 67.843 | 1.00 | 0.00 | C |
| 11 | ATOM | 11 | N2 | G | A | 13 | -21.261 | 2.722 | 66.554 | 1.00 | 0.00 | N |
| 12 | ATOM | 12 | N3 | G | A | 13 | -20.776 | 1.287 | 68.278 | 1.00 | 0.00 | N |
| 13 | ATOM | 13 | C4 | G | A | 13 | -20.504 | 1.257 | 69.600 | 1.00 | 0.00 | C |
| 14 | ATOM | 14 | C5 | G | A | 13 | -20.415 | 2.308 | 70.485 | 1.00 | 0.00 | C |
| 15 | ATOM | 15 | C6 | G | A | 13 | -20.618 | 3.625 | 70.002 | 1.00 | 0.00 | C |
| 16 | ATOM | 16 | O6 | G | A | 13 | -20.562 | 4.691 | 70.626 | 1.00 | 0.00 | O |
| 17 | ATOM | 17 | N7 | G | A | 13 | -20.116 | 1.855 | 71.763 | 1.00 | 0.00 | N |
| 18 | ATOM | 18 | C8 | G | A | 13 | -20.026 | 0.559 | 71.634 | 1.00 | 0.00 | C |
| 19 | ATOM | 19 | N9 | G | A | 13 | -20.245 | 0.128 | 70.345 | 1.00 | 0.00 | N |
| 20 | ATOM | 20 | RPB1 | G | A | 13 | -23.581 | 0.923 | 71.444 | 1.00 | 0.00 | Z |
| 21 | ATOM | 21 | RPB2 | G | A | 13 | -24.021 | 2.216 | 69.838 | 1.00 | 0.00 | Z |

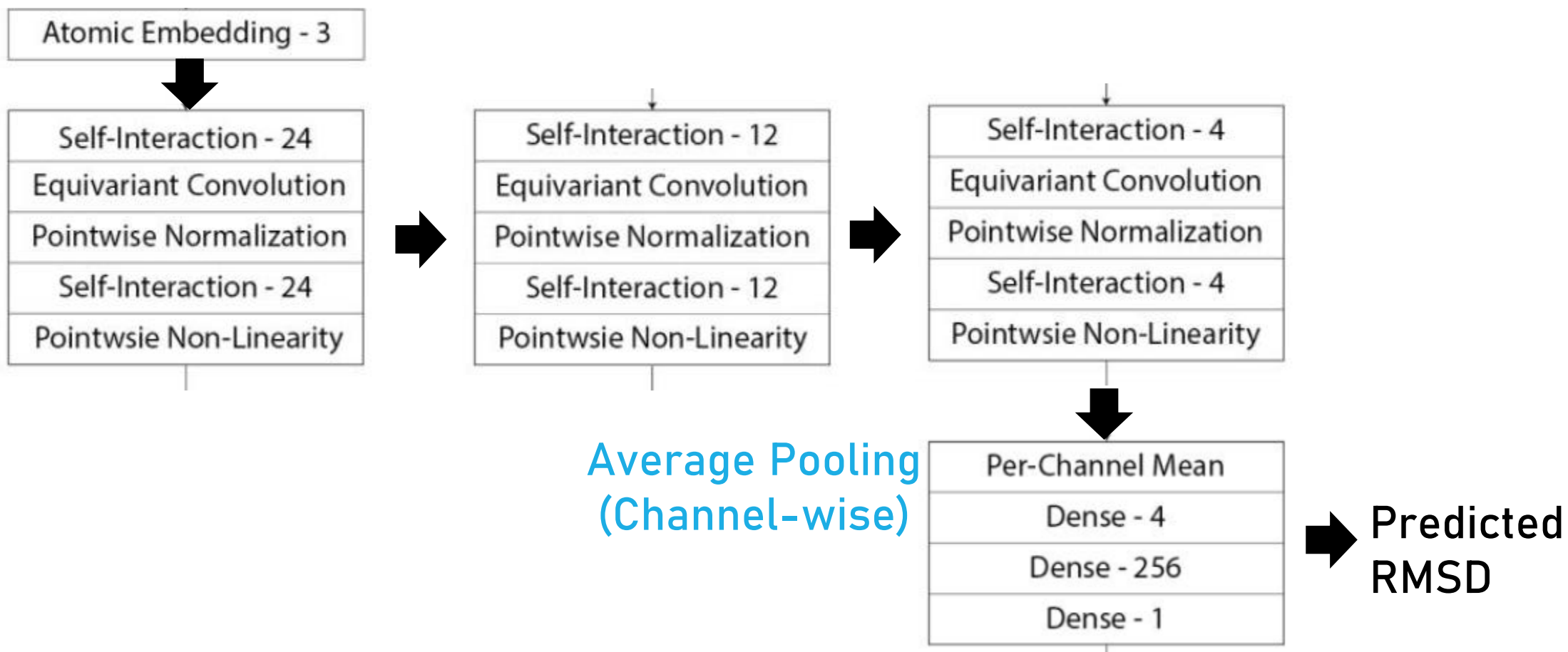
Base

x, y, z-coordinates

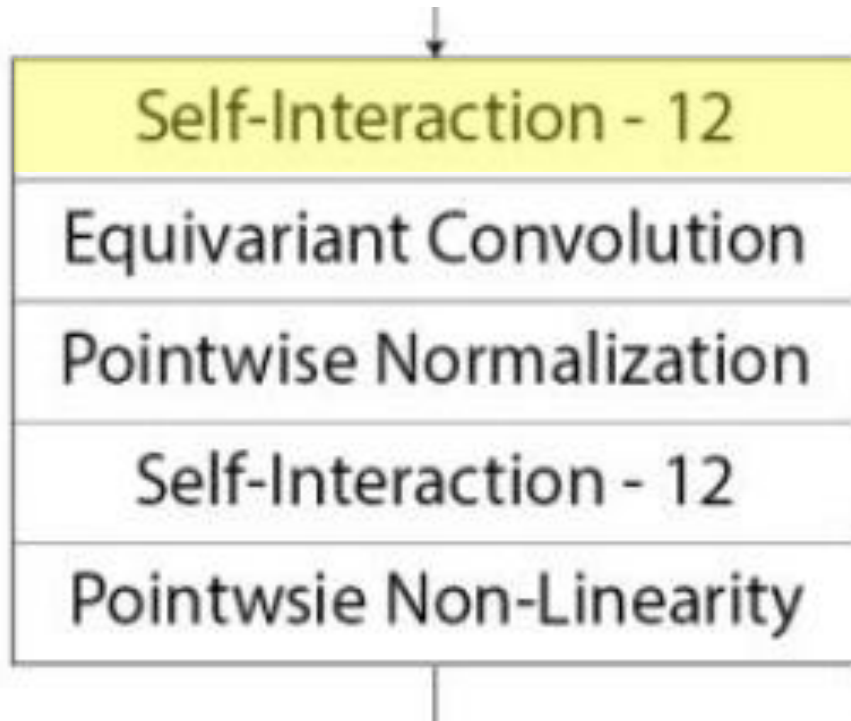
Element

5. ARES Architecture in Detail

One-hot representation
of atoms: C/N/O

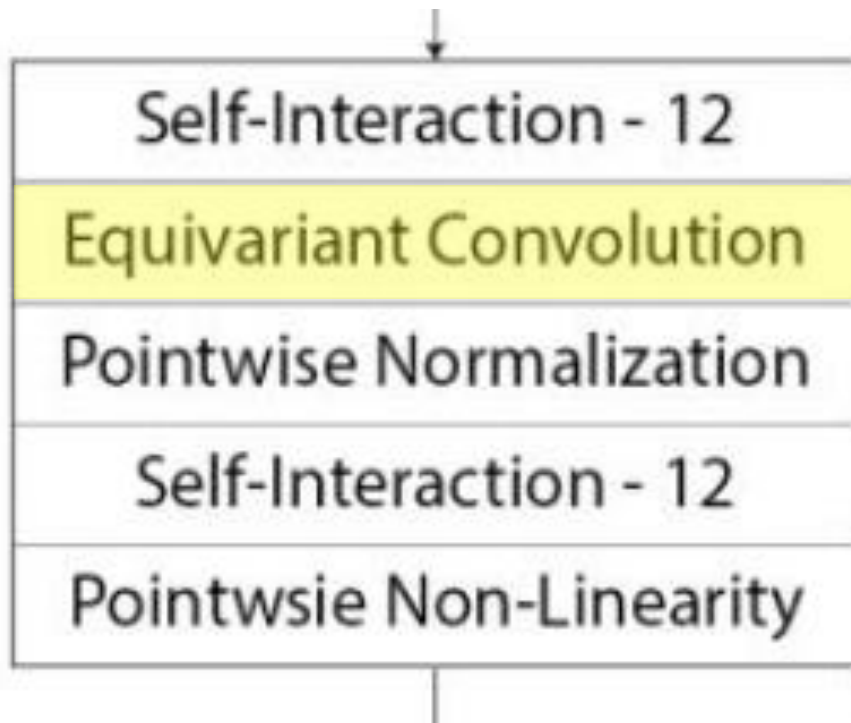


5. ARES Architecture in Detail



- Mixes information across channels.
- Analogous to 1x1 convolutions.

5. Equivariant Point Convolution is the Key of ARES

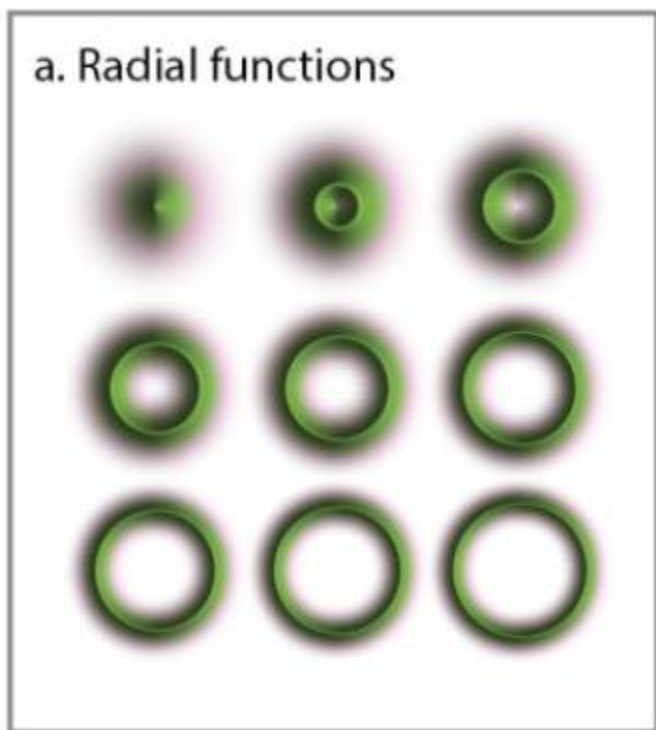


- Neighboring atom's information is shared with the source atom.
- Captures geometric information.
- Rotation and translation equivariant.

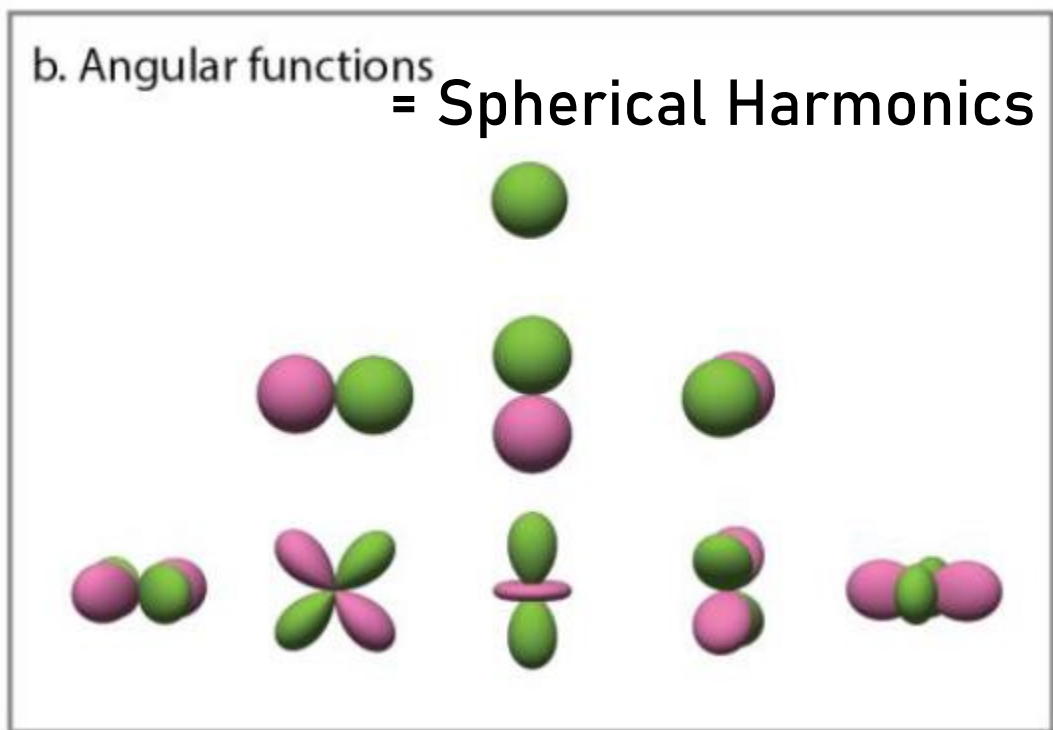
5. ARES Filter is a Product of Radial and Angular Functions

Distance relationships
between atoms

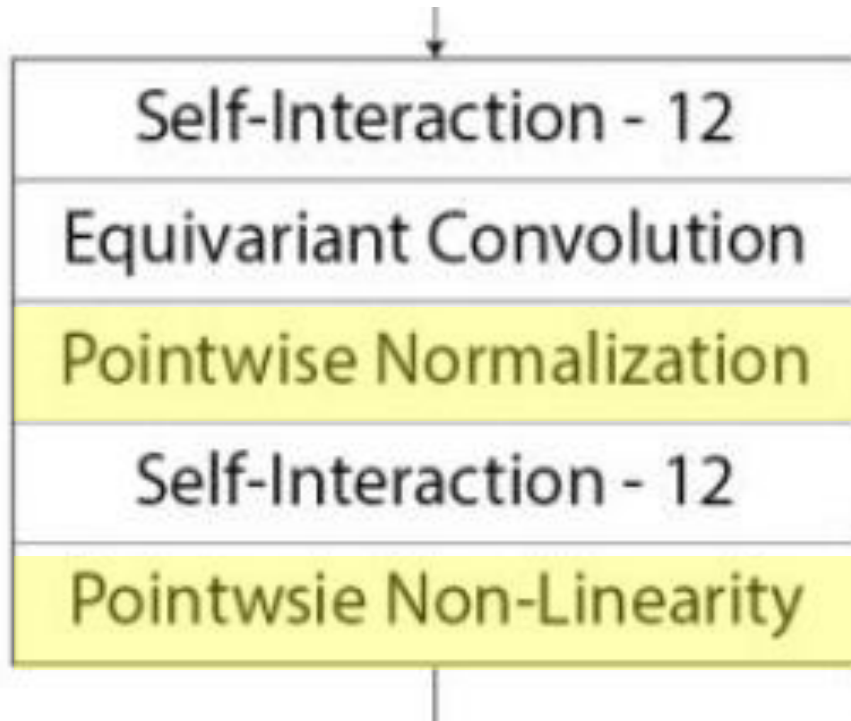
B.



Orientations between atoms



5. ARES Architecture in Detail



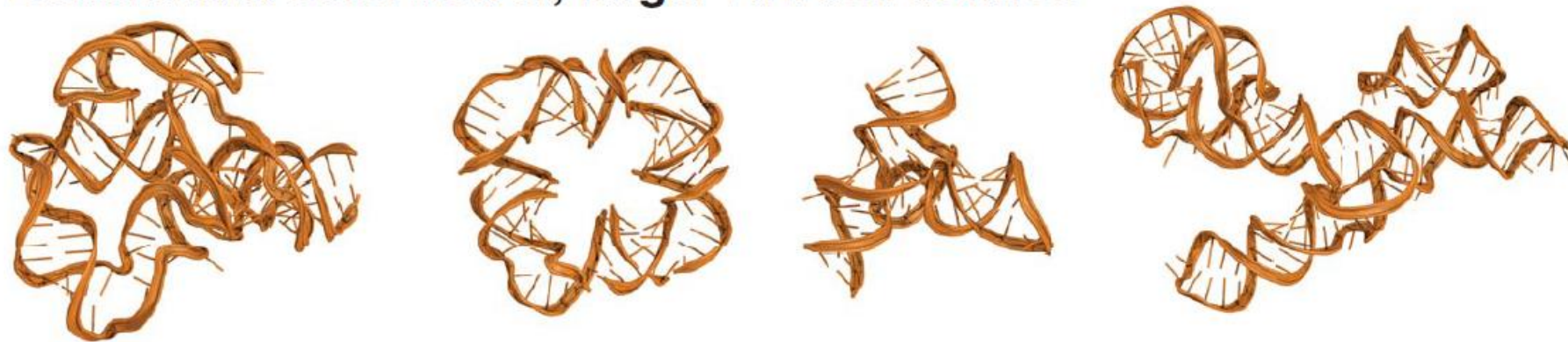
- Point-wise and Feature-wise Norm
- SoftPlus Activation Function

6. Training and Benchmark sets

C Training set: 18 older, smaller RNA structures

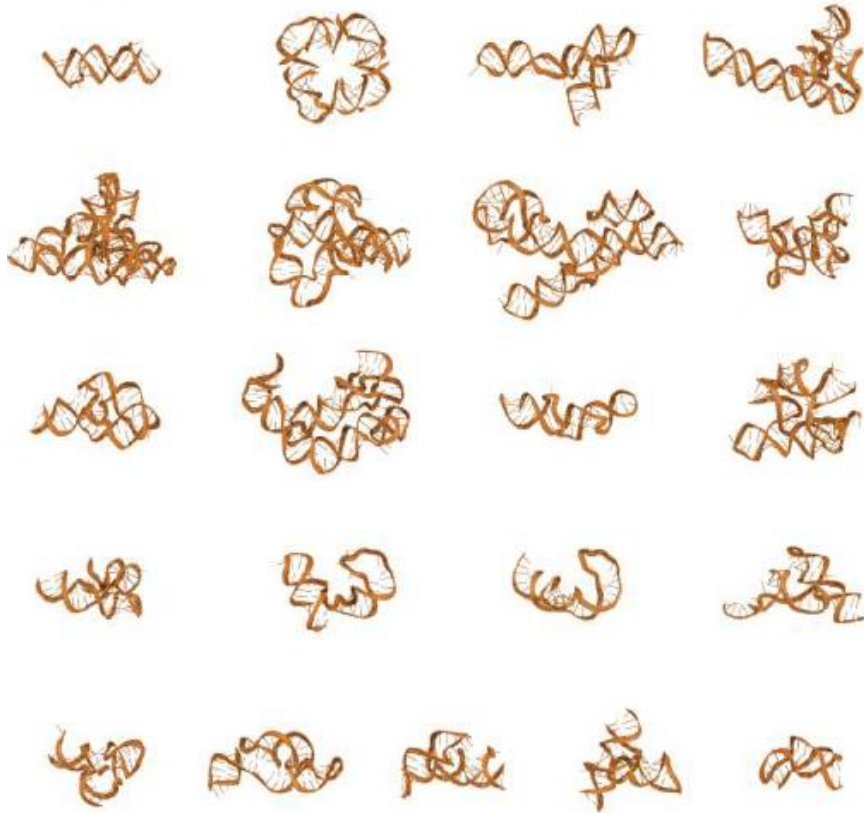


D Benchmark sets: newer, larger RNA structures

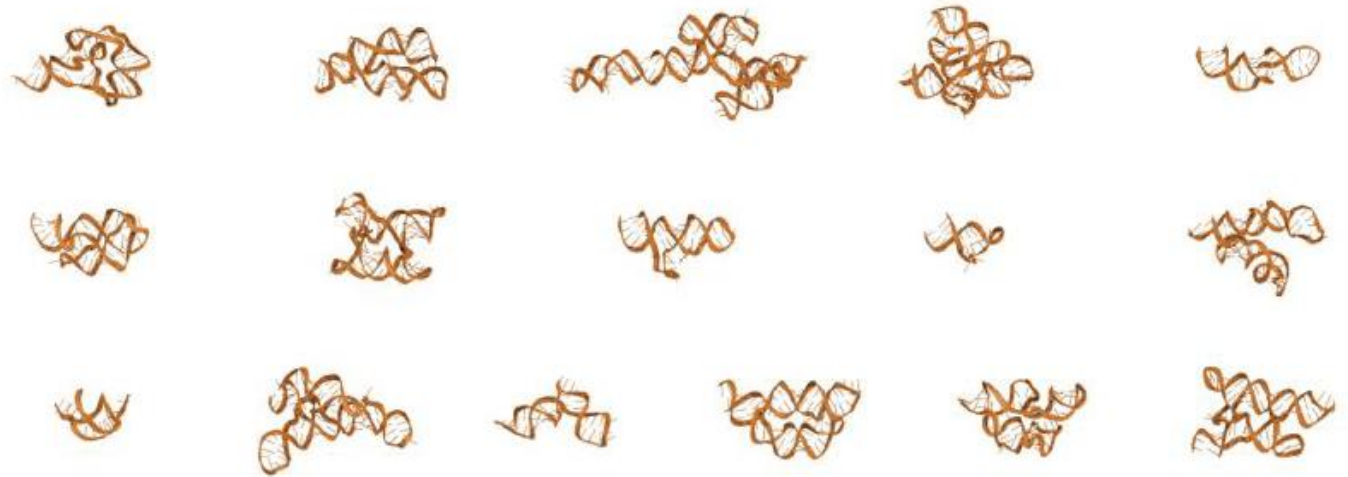


6. Training and Benchmark sets

B. Benchmark 1



C. Benchmark 2



- Benchmark 1: 21 RNA structures used in RNA-Puzzles blind RNA structure prediction challenges.
- Benchmark 2: 16 High-resolution RNA structures from PDBs, published in 2008 or later.

6. Training Details

- **Optimizer:** Adam
- **Loss Function:** Huber
 - Applied to $|(\text{predicted RMSD}) - (\text{true RMSD})|$
- Trained on 18 RNAs.
 - 14 RNAs: training parameters
 - 4 RNAs : hyperparameter optimization.
- 1,000 structural models generated by FARFAR2 were randomly selected and used for each RNA.
- **Batch Size:** 16
- **Learning Rate:** 0.01
- **Number of Epoch:** 1

Background



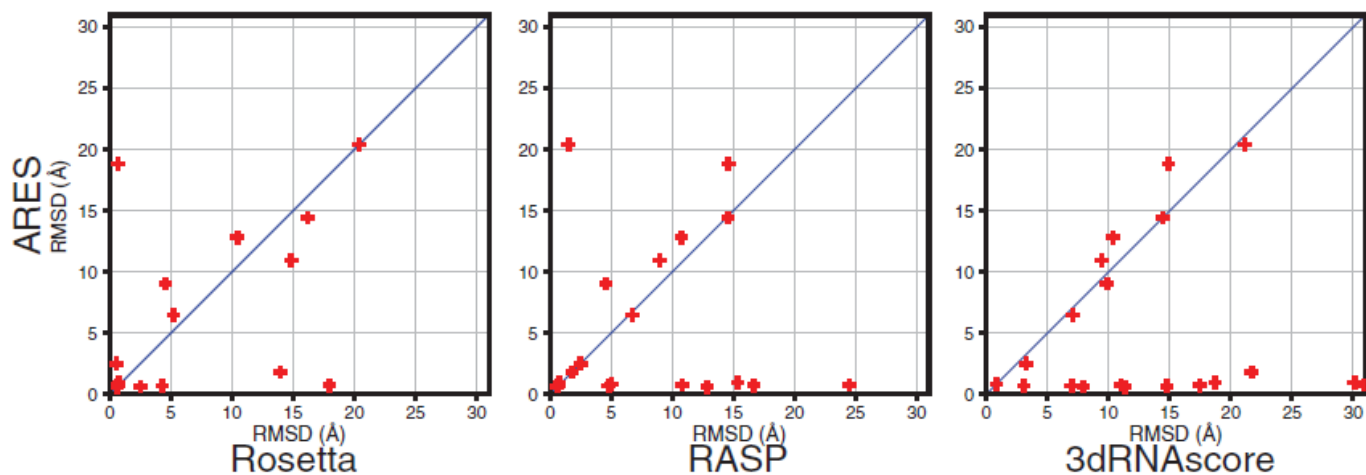
**Architecture
& Training**



Benchmark

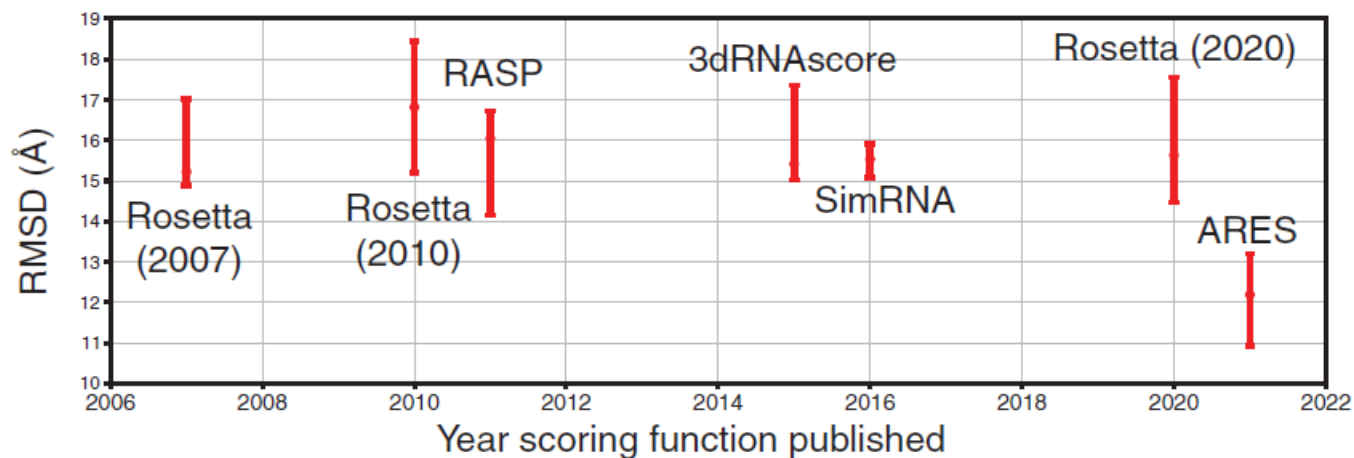
7. ARES Outperforms Other RNA Scorers

A RMSD of best-scoring structural model



Benchmark 1:
near-native models added

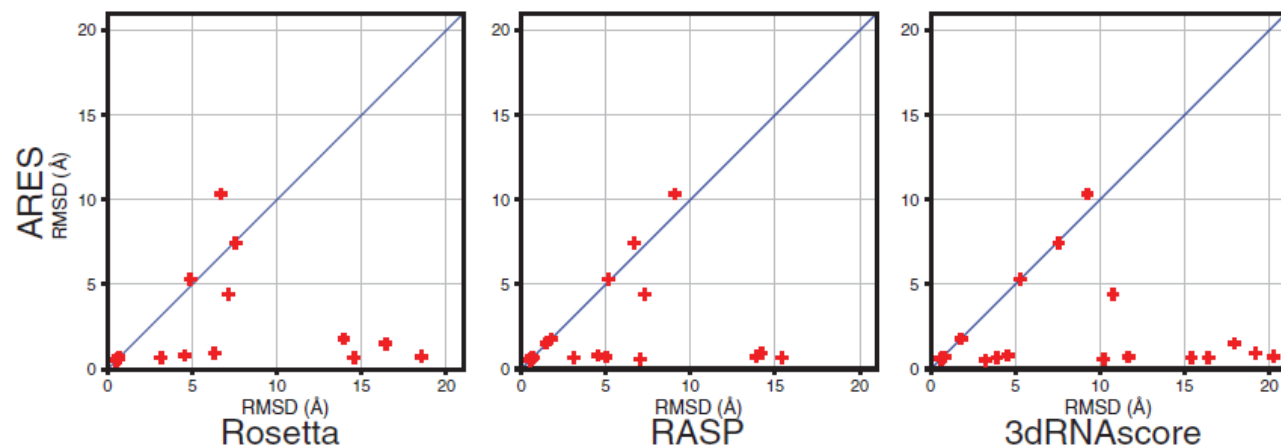
D RMSD of best-scoring structural model



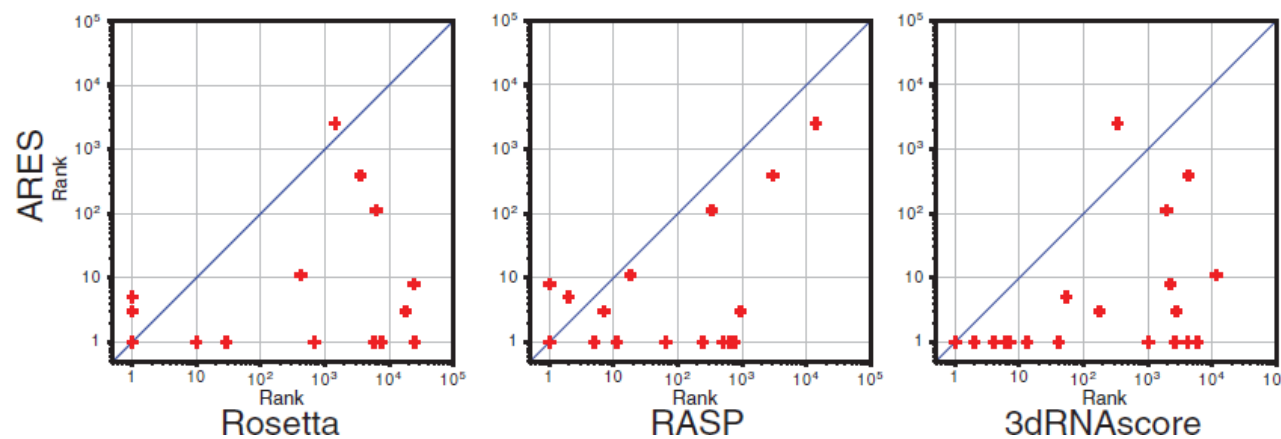
Benchmark 2:
prior information ignored

7. ARES Outperforms Other RNA Scorers

B Lowest RMSD among 10 best-scoring structural models



C Rank of best-scoring near-native structural model



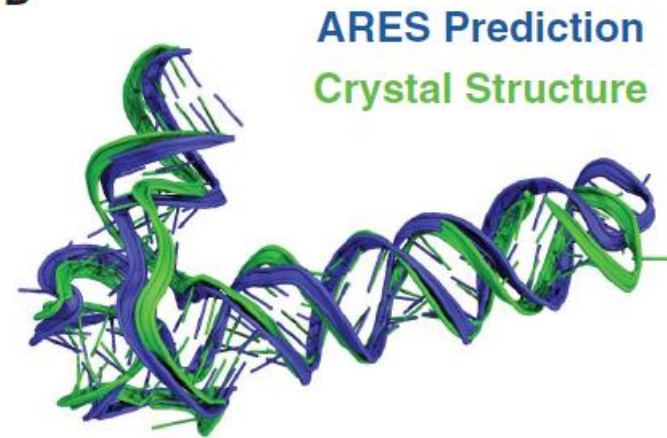
Benchmark 1:
near-native models added

7. ARES Excels in Blind RNA Structure Prediction

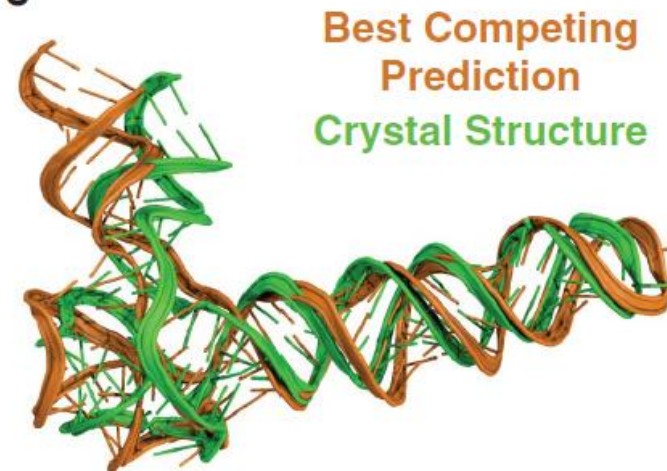
A Blind prediction accuracy (RMSD, Å)

| Method | RNA | | | |
|-------------|------------|-------------|------------|-------------|
| | A | B | C | D |
| ARES | 4.8 | 12.5 | 9.5 | 14.5 |
| Adamiak | 9.8 | 18.7 | 19.1 | 18.2 |
| Bujnicki | 9.8 | 14.0 | 15.6 | 20.0 |
| Chen | 11.0 | 18.1 | 11.7 | 32.8 |
| Ding | 19.1 | 17.4 | — | 34.3 |
| Das (Human) | 13.6 | 13.3 | 10.1 | 28.8 |
| iFoldRNA | 10.3 | 23.5 | 53.3 | 22.4 |
| RNAComposer | 10.2 | 19.0 | 14.1 | 19.6 |
| Rosetta | 7.7 | 14.3 | 10.1 | 22.2 |
| SimRNA | 13.7 | 16.2 | 42.2 | 22.2 |
| Xiao | 15.4 | 20.6 | 27.2 | 29.4 |

B



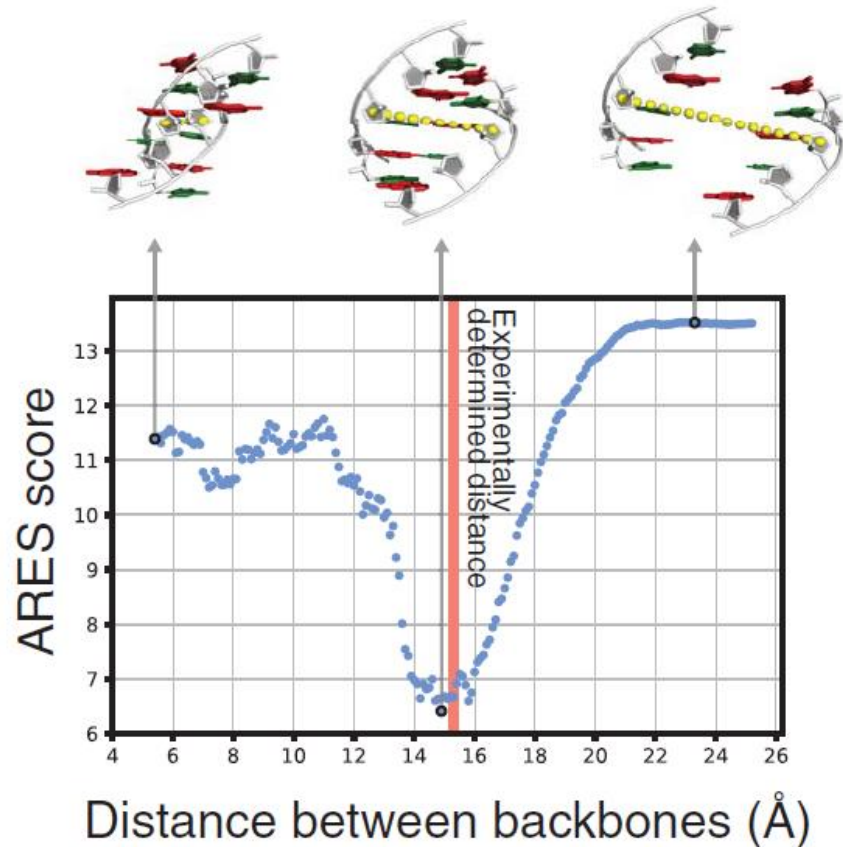
C



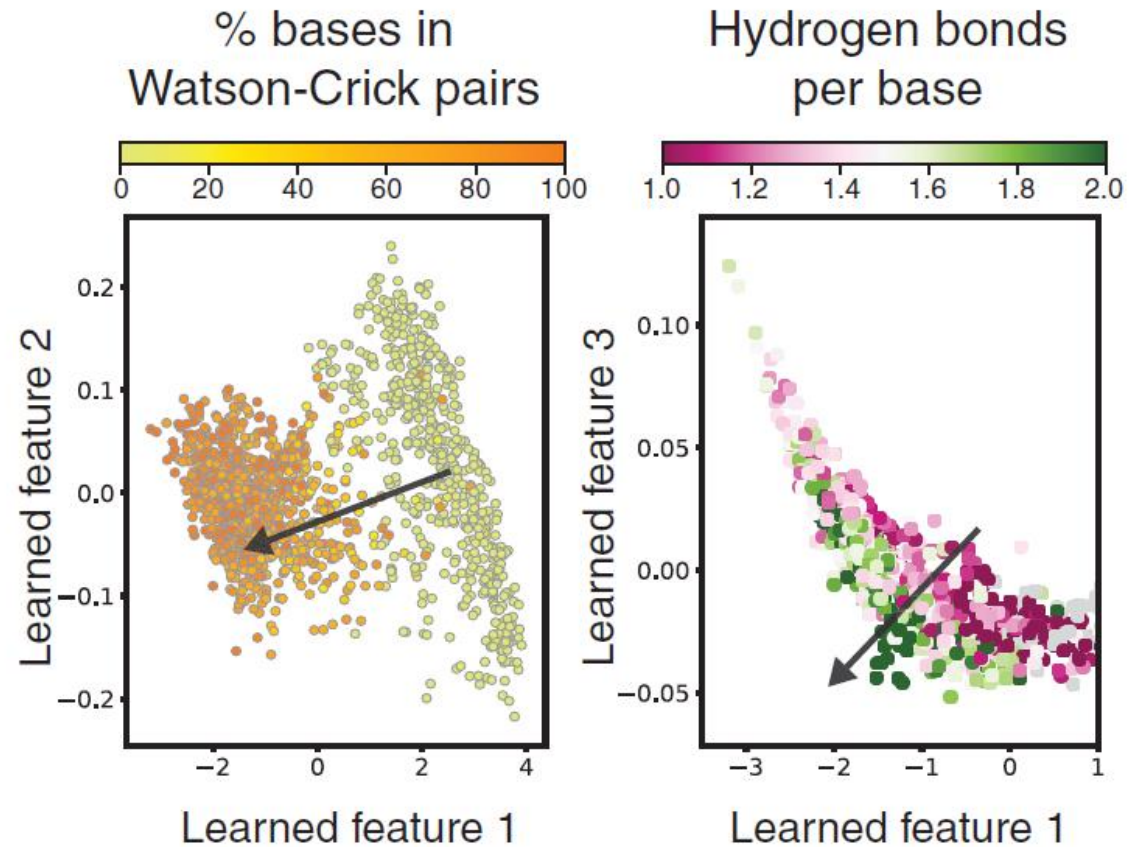
A: Adenovirus VA-I RNA
B-D: Bacterial tRNAs

8. ARES Identifies Key Structural Characteristics

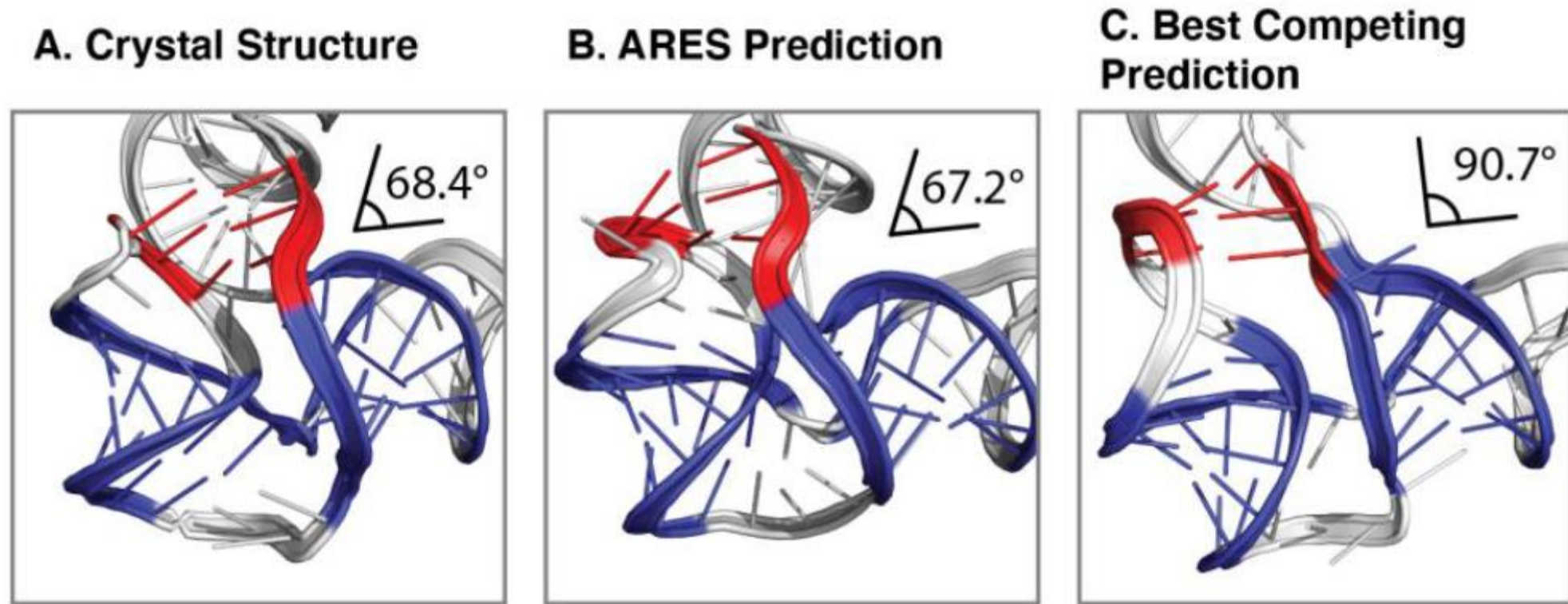
A ARES learns helix width for optimal base pairing



B ARES learns to identify key RNA characteristics



8. ARES Recovers of a Non-trained Structural Motif



An unusual 3-way junction , whose geometry is defied by a **pseudoknot**.

8. ARES Identifies RNA Tertiary Interactions

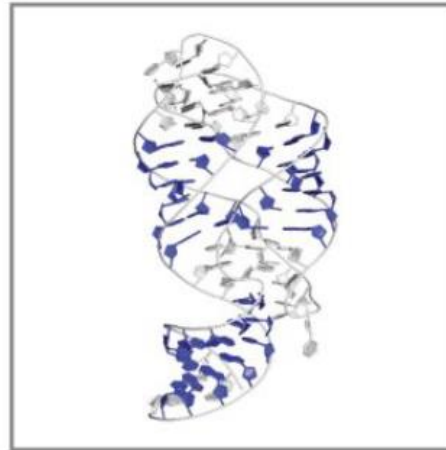
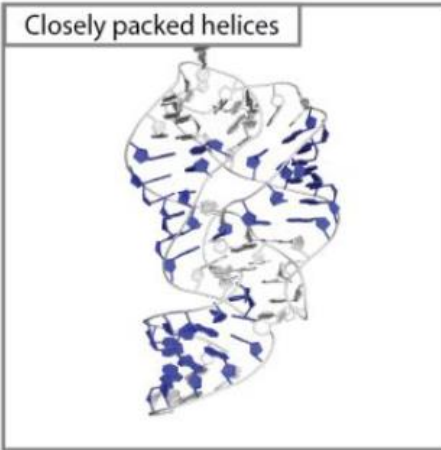
Closely packed tertiary structure results

ARES Prediction

Crystal Structure

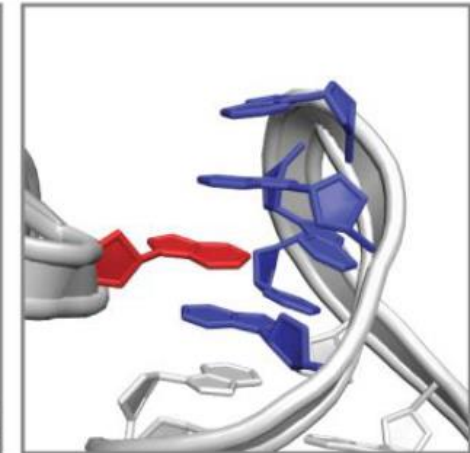
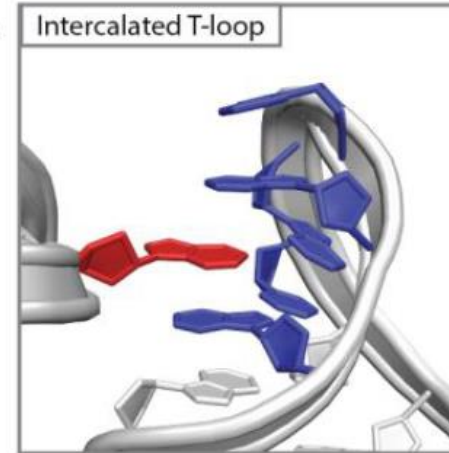
A.

Closely packed helices



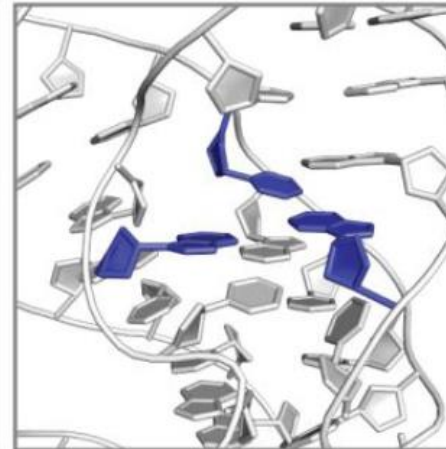
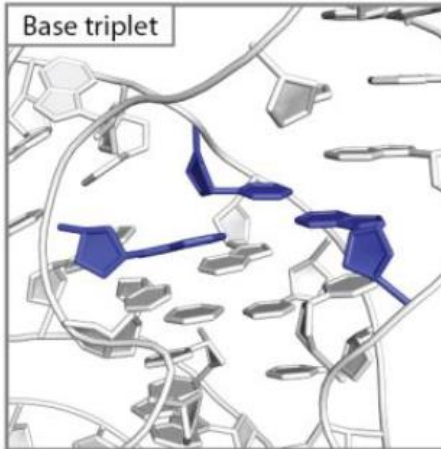
C.

Intercalated T-loop



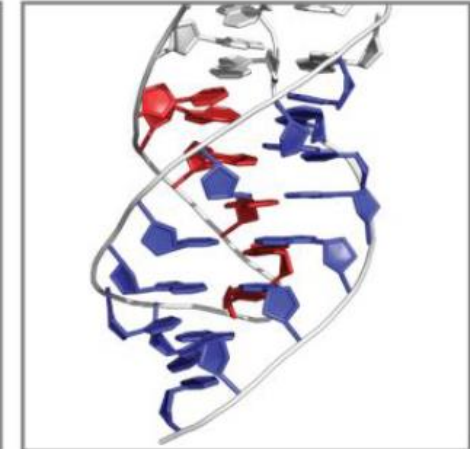
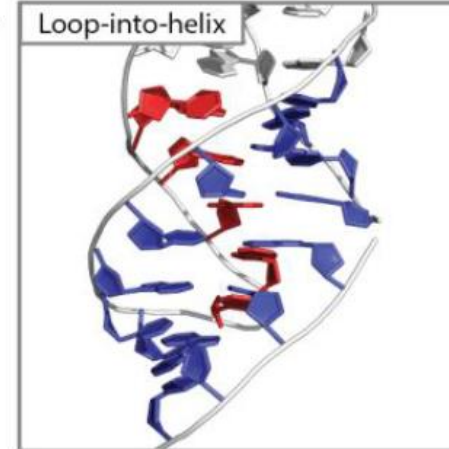
B.

Base triplet



D.

Loop-into-helix



9. Conclusion

1. ARES learns **atomic structures** of RNA and scores the accuracy of predicted structures.
2. ARES network **has translational and rotational equivariance**, which allows ARES to properly capture geometric information.
3. ARES **outperforms** competing RNA structure scorers.
4. ARES can be used to identify structural characteristics, motifs, and tertiary interactions.

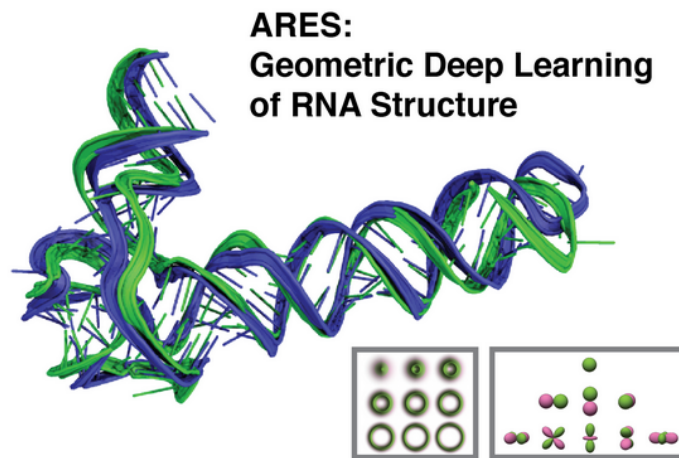
10. Further Research Directions

1. ARES relies on external programs to generate candidate structural models.
2. ARES can be **incorporated into folding programs** to guide candidate model generation.
3. ARES can be trained to predict structures of proteins, DNA, or Drugs.
4. ARES can be used outside of structure prediction.

References

1. Townshend, R. J., Eismann, S., Watkins, A. M., Rangan, R., Karelina, M., Das, R., & Dror, R. O. (2021). Geometric deep learning of RNA structure. *Science*, 373(6558), 1047-1051.
2. Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K., & Riley, P. (2018). Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. arXiv preprint arXiv:1802.08219.
3. Singh, J., Hanson, J., Paliwal, K., & Zhou, Y. (2019). RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nature communications*, 10(1), 1-13.
4. Watkins, A. M., Rangan, R., & Das, R. (2020). FARFAR2: Improved de novo Rosetta prediction of complex global RNA folds. *Structure*, 28(8), 963-976.
5. Andrzejewska, A., Zawadzka, M., & Pachulska-Wieczorek, K. (2020). On the way to understanding the interplay between the RNA structure and functions in cells: A genome-wide perspective. *International Journal of Molecular Sciences*, 21(18), 6770.
6. Deshpande, A. (2016). A Beginner's Guide To Understanding Convolutional Neural Networks, (<https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks>)
7. Smidt, T. (2018), Machine Learning for Molecules and Materials, NeurIPS Workshop

How to Use: Pre-trained ARES Webserver



To make a prediction with ARES, submit a Rosetta 'silent file' or a PDB file containing candidate structural models (not both!), and ARES will output a single score (lower is better) for each model in the submitted file. For more details on ARES, please check out our [paper](#)!

Server ready with 9152 job(s) in queue.

Email for results:

[either] Rosetta 'silent file':

No file selected.

[or] PDB file:

No file selected.

Upload progress:

<http://drorlab.stanford.edu/ares.html>



<https://colab.research.google.com/drive/1yczwMBkQ-MOBw00ilqukW0Puzav1tK5K?usp=sharing>

Geometric Deep Learning of RNA Structure

Thank you for your attention!