# lociPARSE: a locality-aware invariant point attention model for scoring RNA 3D structures

Sumit Tarafder and Debswapna Bhattacharya

Department of Computer Science, Virginia Tech, USA

## Background

- Knowledge-based statistical potentials and deep-learning for RNA quality estimation
- Deep-learning based methods rely on superposition-dependent RMSD
- Scoring should be invariant to relative motion among structural motifs

### Novelty

- First method targeting lDDT as ground truth
- Locality-aware invariant point attention based on atomic coordinates
- First deep learning method to estimate RNA local and global quality in a superposition free way
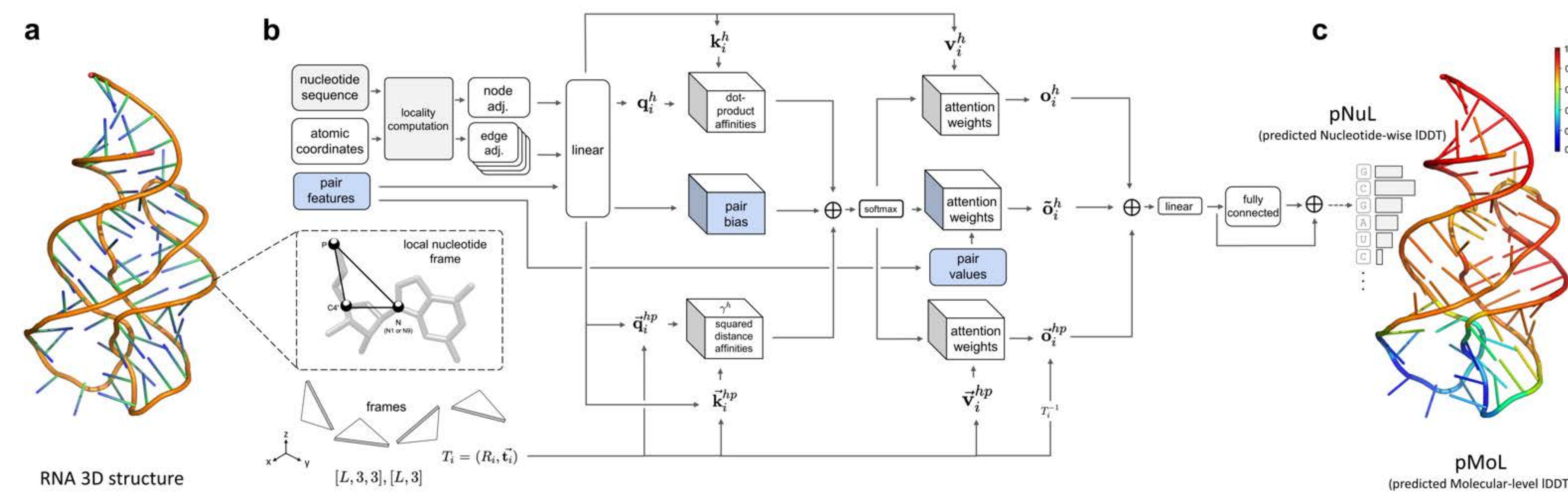
### Workflow

1. Extract features from any input RNA 3D structure
2. Generate per nucleotide frame from P, C4′ and N coordinates
3. Aggregate features based on K-nearest neighbors
4. Compute invariant point attention operating on frames
5. Predict quality score (0-1]
   - pMoL = predicted molecular-level lDDT
   - pNuL = predicted nucleotide-wise lDDT

### Source

## Overview



## Performance on independent test set

**Table 1.** Performance on 30 independent RNA targets in terms of lDDT as ground truth, sorted in decreasing order of global Pearson's $r$. Best values are highlighted in bold.

| Method | Global | | | | | Per-target average | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $r$ ↑ | $\rho$ ↑ | $\tau$ ↑ | Diff ↓ | ROC ↑ | $r$ ↑ | $\rho$ ↑ | $\tau$ ↑ | Loss ↓ |
| lociPARSE | **0.67** | **0.71** | **0.55** | **0.06** | **0.91** | **0.77** | **0.75** | **0.64** | 0.06 |
| rsRNASP | 0.5 | 0.5 | 0.36 | 0.11 | 0.79 | 0.75 | 0.69 | 0.55 | 0.06 |
| RASP | 0.46 | 0.5 | 0.36 | 0.16 | 0.78 | 0.72 | 0.68 | 0.58 | 0.08 |
| ARES | 0.43 | 0.48 | 0.34 | 0.55 | 0.8 | 0.72 | 0.7 | 0.58 | 0.06 |
| DFIRE-RNA | 0.33 | 0.32 | 0.22 | 0.19 | 0.69 | 0.75 | 0.7 | 0.56 | **0.05** |
| cgRNASP | 0.27 | 0.19 | 0.13 | 0.15 | 0.60 | 0.12 | 0.07 | 0.07 | 0.07 |
| RNA3DCNN | 0.19 | 0.17 | 0.12 | 0.64 | 0.58 | 0.47 | 0.36 | 0.27 | 0.06 |

## Performance on CASP15

**Table 2.** Performance on CASP15 RNA targets in terms of lDDT as ground truth, sorted in decreasing order of global Pearson's $r$. Best values are highlighted in bold.

| Method | Global | | | | | Per-target average | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $r$ ↑ | $\rho$ ↑ | $\tau$ ↑ | Diff ↓ | ROC ↑ | $r$ ↑ | $\rho$ ↑ | $\tau$ ↑ | Loss ↓ |
| lociPARSE | **0.74** | **0.72** | **0.55** | 0.12 | **0.96** | **0.73** | **0.66** | **0.52** | **0.07** |
| ARES | 0.33 | 0.35 | 0.25 | 0.27 | 0.78 | 0.55 | 0.53 | 0.42 | 0.13 |
| RASP | 0.33 | 0.42 | 0.3 | 0.17 | 0.80 | 0.55 | 0.54 | 0.42 | 0.17 |
| rsRNASP | 0.31 | 0.41 | 0.3 | **0.11** | 0.83 | 0.66 | 0.61 | 0.47 | 0.11 |
| DFIRE-RNA | 0.29 | 0.3 | 0.22 | 0.39 | 0.77 | 0.69 | 0.6 | 0.46 | 0.16 |
| cgRNASP | 0.29 | 0.32 | 0.24 | 0.12 | 0.80 | 0.59 | 0.5 | 0.37 | 0.11 |
| RNA3DCNN | 0 | -0.01 | -0.01 | 0.52 | 0.35 | 0.5 | 0.42 | 0.31 | 0.09 |

## Case study
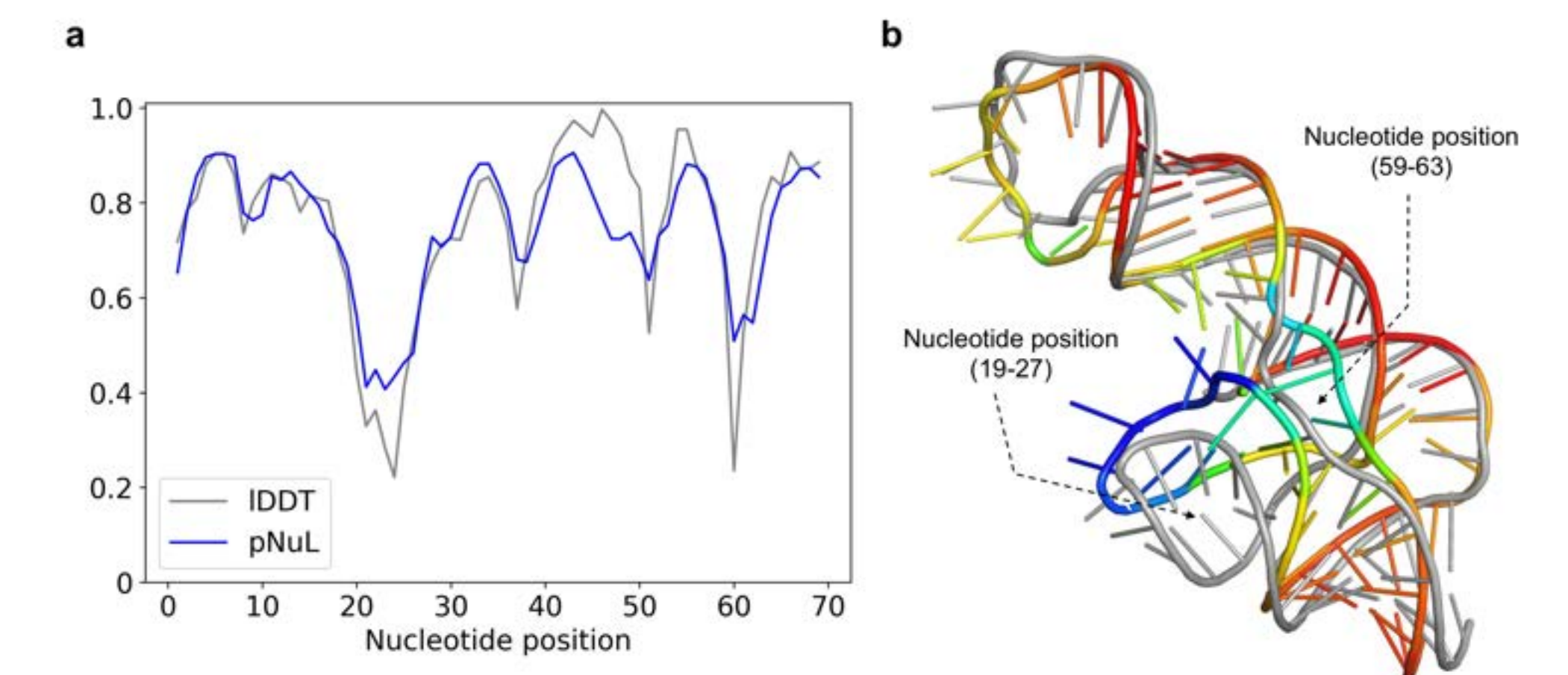


Per nucleotide prediction accuracy on 7QR3

## Local quality prediction

**Table 3.** Local nucleotide-wise scoring performance on CASP15 set. Best values are highlighted in bold.

| Method | Local nucleotide-wise | | | |
|---|---|---|---|---|
| | $r$ ↑ | $\rho$ ↑ | $\tau$ ↑ | Diff ↓ |
| lociPARSE | **0.72** | **0.73** | **0.53** | **0.15** |
| RNA3DCNN | 0.17 | 0.19 | 0.13 | 0.24 |

## Ranking ability
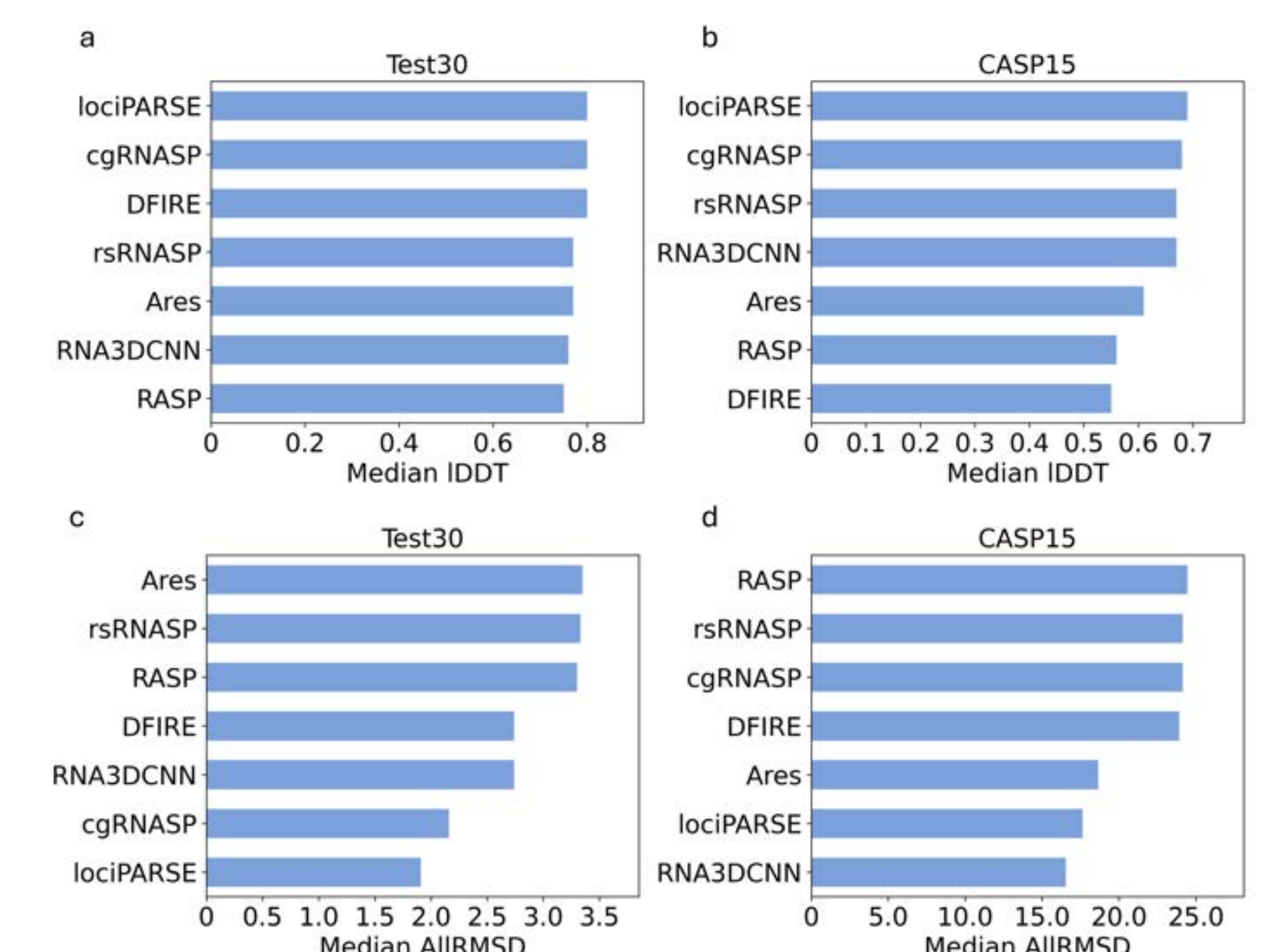


Median scores of top ranked structure

## Conclusion

- Can be extended for RMSD or INF
- Multi-reference lDDT for enhanced expressivity