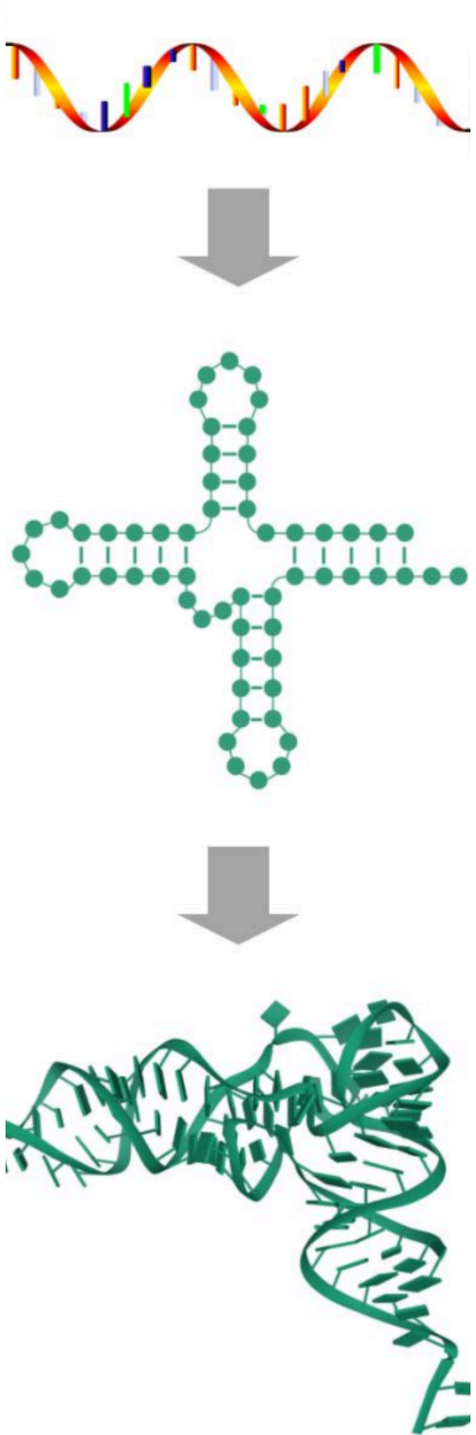
RSSP - Tool to Predict RNA Secondary Structures Using Multiple Computational Approaches

Jakub Binda

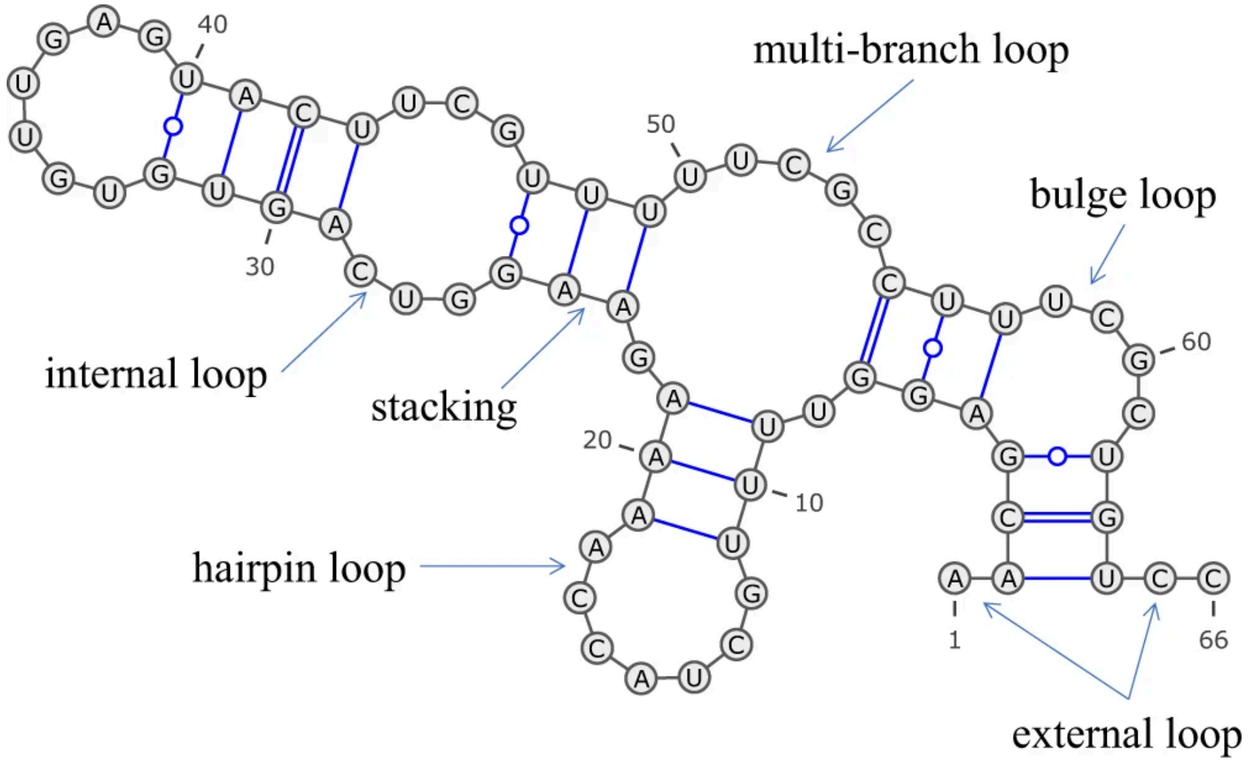
Konstanty Kraszewski

Anastazja Ramanczanka

Michalina Wysocka



RNA molecules fold into characteristic secondary and tertiary structures that account for their diverse functional activities.



Nat Commun 12, 941 (2021). https://doi.org/10.1038/s41467-021-21194-4

Secondary structure of RNA can be determined from atomic coordinates obtained by **X-ray crystallography**.

Limitations:

- high experimental costs,
- resolution limits on measurements of RNA.

Solution: perform computational prediction.

Prediction Approaches

- Dynamic programming and thermodynamic models.

  Mfold/UNAfold, RNAfold, RNAstructure
- Machine learning models.

  CONTRAfold, ContextFold
- Hybrid methods that combine thermodynamic and ML-based approaches.

  SimFold, MXfold

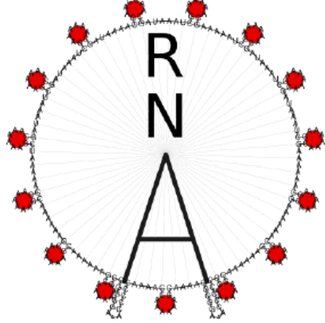
Program name: RSSP (RNA Secondary Structure Predictors)

Objective: To produce accurate RNA secondary structures by integrating predictions from multiple computational tools, each employing different algorithms and approaches.

github.com/jbindaAI/RSSP

- **RNA Sequence Input:** Load RNA sequences from FASTA file or enter sequences directly in FASTA format.
- **Tools Integration:** Seamlessly run MXFold2, KnotFold, RNAFold and RNAstructure on the input sequences.
- **Dot-Bracket Notation:** Get the predictions in a dot-bracket format for easy analysis.
- **Result Visualization:** Generate visualizations to help interpret the predicted RNA secondary structures.
- **User-friendly Interface:** Simple and intuitive interface for running predictions and viewing results.

RNAfold (ViennaRNA package)



Employs dynamic programming techniques to find the most thermodynamically stable structure for a given RNA sequence.

```
> test
CUACGGCGCGGCGCCCUUGGCGA
```

Compute the best (MFE) structure for this sequence

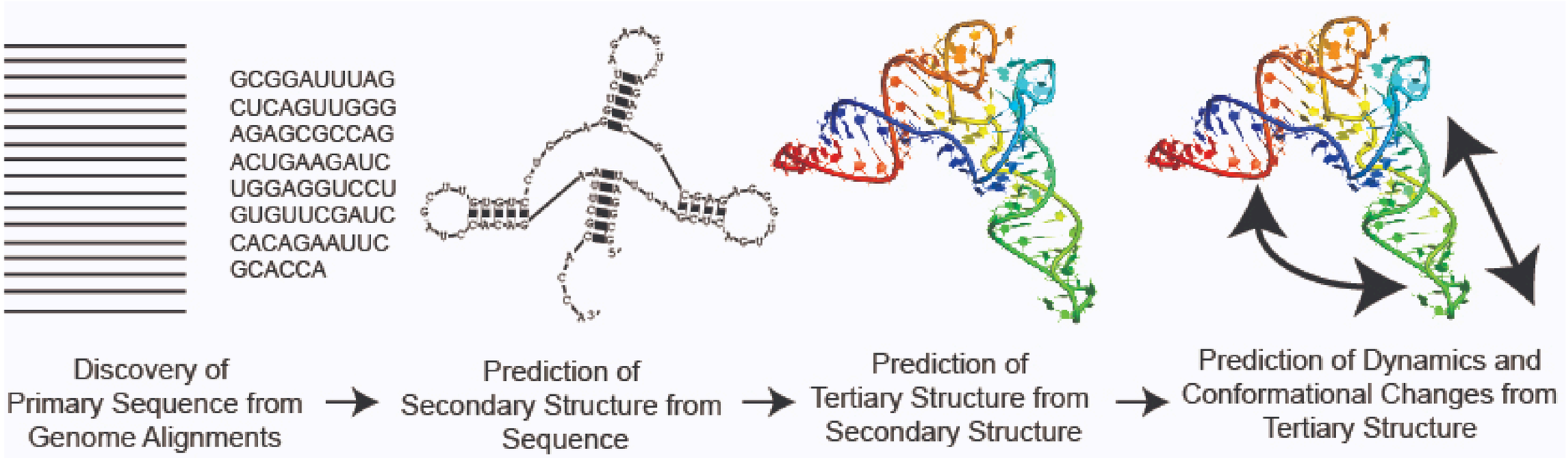
```
$ RNAfold < test.seq
CUACGGCGCGGCGCCCUUGGCGA
...........((((...)))). ( -5.00)
```

dot-bracket notation -> **<- minimum free energy of structure**

github.com/ViennaRNA/ViennaRNA

RNAstructure

- Automation of modeling of RNA structure and function from genome sequence to 3D structure.
- Developed by the Mathews lab.



https://rna.urmc.rochester.edu/index.html

MXFold2

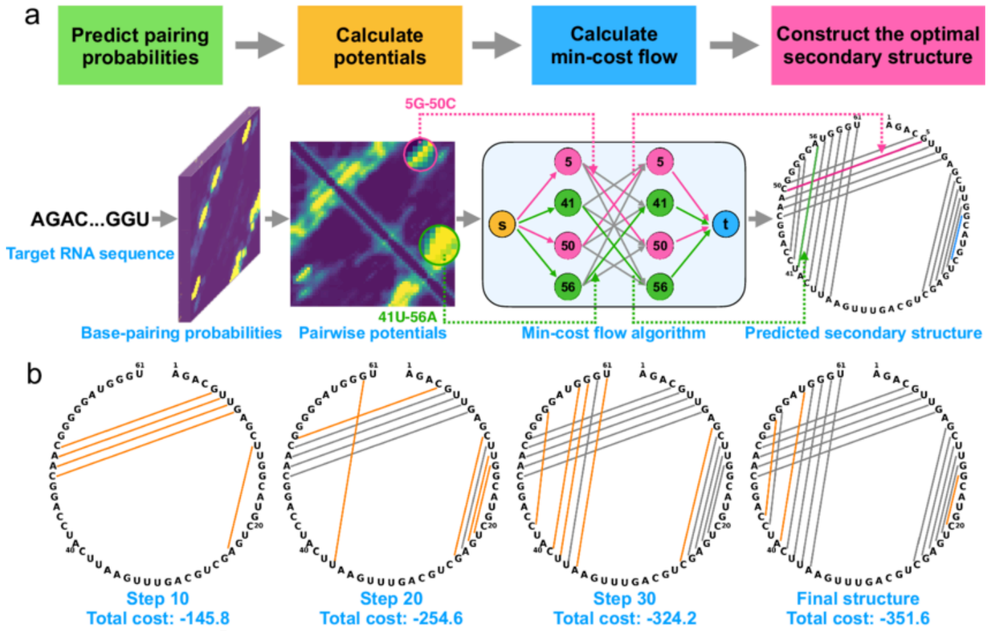
- Deep Learning based method (CNN+BiLSTM).
- Integrated with thermodynamics approach.
- DNN predicts folding scores for each base pair.
- Folding scores are used by a classic nearest neighbor RNA model.

https://github.com/mxfold/mxfold2.git

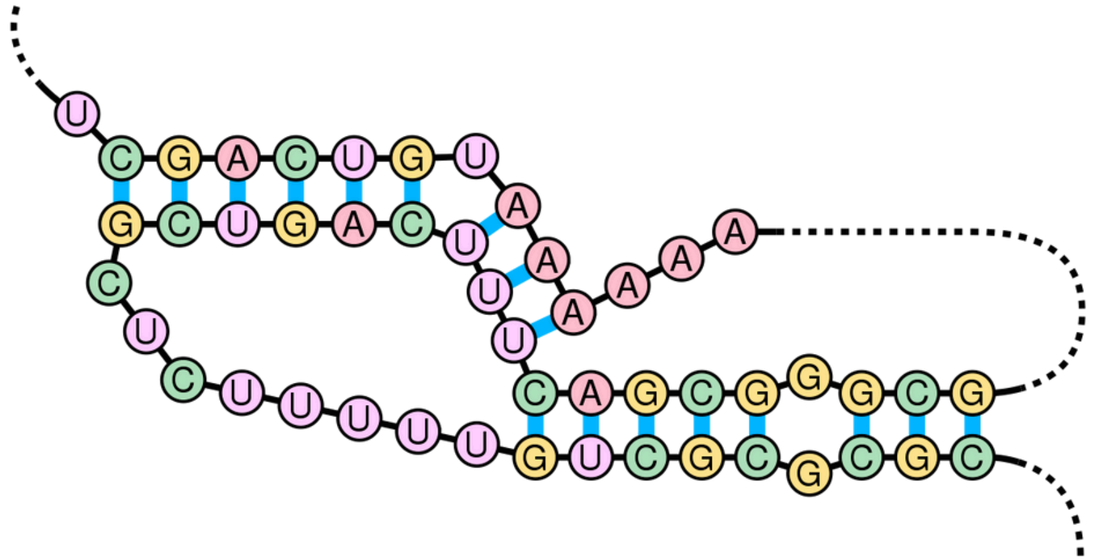
KnotFold

ML approach to predict RNA structures including pseudoknots, by finding the secondary structure with the lowest potential.

It uses attention-based neural network, avoiding inaccuracy of hand-crafted energy functions.



DOI:10.1038/s42003-024-05952-w



wikipedia.org/wiki/Pseudoknot

Demo version of the web app

RSSP

FOLD RNA

Source code

Version control


Frontend&Backend

FastAPI

HTML

JS

B

Future Work and Improvements

- Incorporate additional prediction tools employing different approaches.
- Add a description for the resulting loops on the visualization.
- Define and implement metrics to compare the performance of different RNA secondary structure prediction tools.

Conclusions

Accurate secondary structure prediction is crucial for understanding RNA function, interactions, and mechanisms.

By combining the strengths of various tools we can achieve more accurate and reliable predictions.