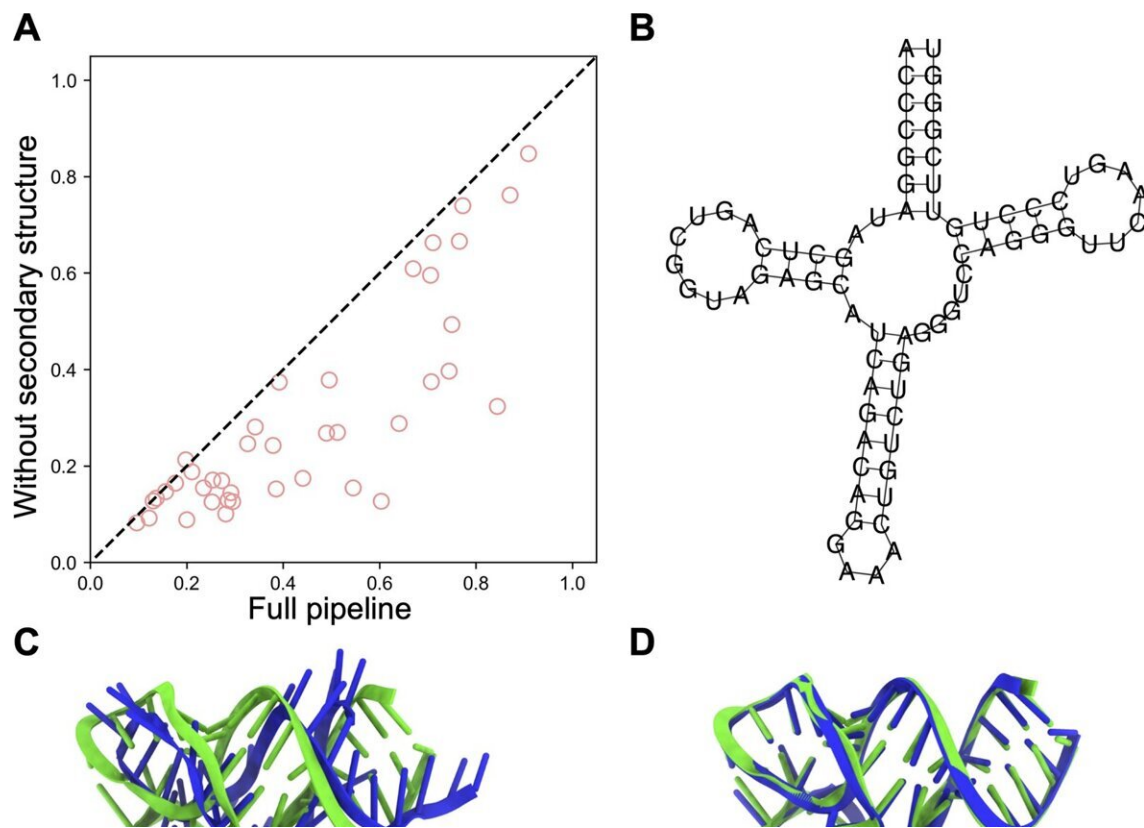


Novel AI-based approach for more accurate RNA 3D structure prediction

December 21 2023



Secondary structure feature improves performance. Credit: *Nature Communications* (2023). DOI: 10.1038/s41467-023-41303-9

A research team from the Cancer Science Institute of Singapore (CSI Singapore) at the National University of Singapore (NUS) has

successfully harnessed artificial intelligence (AI) and deep-learning techniques to model atomic-level RNA 3D structures from primary RNA sequences. Called DRfold, this novel AI-based method improves the accuracy of RNA models by more than 70%, compared to traditional approaches.

The team, which is led by Professor Zhang Yang from CSI Singapore and NUS School of Computing, published their findings in [*Nature Communications*](#) on 16 September 2023.

RNAs are large biomolecules consisting of a single chain of nucleotides, which derive their sequence order from double-stranded DNA molecules during transcription. RNAs are widely known for their role in transcription and translation processes, which facilitates the transfer of gene information embodied in DNA sequences into protein amino acid sequences.

In recent years, RNAs have been found to play important roles in regulating various biological processes, hence positioning them as novel drug targets.

It has been estimated that targeting RNAs with small molecules will expand the drug design landscape exponentially, compared to traditional protein-targeted drug discovery. Accordingly, RNA biology and its applications in developing new therapeutics represent a critical emerging field, garnering significant academic and industry investment worldwide.

Predicting RNA structures

Compared to well-folded protein structures, RNA structures and their folds are generally considered less stable due to the relatively shallow energy landscape. Therefore, traditional physics- and statistics-based force fields, which are often error-prone, cannot accurately describe the

elegant and intricate folding interactions of RNAs

Meanwhile, the limited availability of experimental RNA structures in the Protein Data Bank (PDB) further constrains the accuracy of these traditional knowledge-based force fields, which are derived from the statistics of the PDB structures.

To address these challenges, DRfold created two complementary deep-learning network pipelines—one focused on end-to-end learning, and the other on geometrical restraint learning. This innovative approach significantly improved the accuracy of the AI-based force field. The synergistic coupling of these two networks also further enhanced the accuracy of the single neural network-based AI potentials.

The key innovation lies in introducing a deep learning approach for predicting RNA tertiary structure. While traditional methods relied on homologous modeling or physics-based folding simulations, which suffer from the limitation of the force field accuracy, DRfold uses self-attention transformer networks to predict 3D structures from RNA sequences, marking a revolutionary shift in addressing this crucial challenge.

DRfold's new strategy of integrating two parallel and complementary networks built on end-to-end and geometry learnings helps to enhance the accuracy of the potential function and RNA model prediction, making it light, highly flexible, scalable, and hence, the preferred prediction method.

Dr. Li Yang, a Research Scientist at CSI Singapore and first author of this study, said, "Since the biological functions of RNAs depend on the specific tertiary structures, it becomes increasingly important and necessary to determine the 3D structures of RNAs in order to facilitate RNA-based function annotation and drug discovery."

He added, "The golden standard in structural biology, such as using biophysical experiments—X-ray crystallography, Cryogenic Electron Microscopy (Cryo-EM), and Nuclear Magnetic Resonance (NMR) Spectroscopy—to determine RNA structures, are often cost- and labor-intensive, limiting their application to a tiny portion of known RNAs."

"Currently, there are more than 30 million known RNA sequences in the RNA central database, but only less than 500 (or 0.0017%) have experimentally solved structures. This frustratingly leaves more than 99% of RNA targets with no structural information. Hence, our study's core aim is to develop new computational methods capable of predicting high-quality RNA structure models, filling this substantial information gap."

Potential applications in drug design and virtual screening

Prof Zhang, Senior Principal Investigator at CSI Singapore and corresponding author of the study, said, "Our primary goal for this study is to bridge the gap between the scarcity of experimental RNA structures and the increasing demand of the RNA biology field and drug industry. In this regard, high-confident DRfold models can be used as a starting point to guide the RNA drug design and virtual screening, or to help elucidate the biological functions of the RNA molecules in cells."

"Considering the potency and effectiveness of mRNA vaccines in combating pandemics, tools such as DRfold play a crucial role in predicting and optimizing RNA structures and the stability of vaccines. Furthermore, these tools can be used to study the biological functions of RNAs, particularly non-coding RNAs, and design novel RNA experiments using predicted models which follow the sequence-to-structure-to-function paradigm," Prof Zhang added.

The group has opened the source codes of DRfold to the public community via their webpage: <https://zhanggroup.org/DRfold>. Its high scalability and open-source framework render it incredibly flexible and applicable for solving other related problems, such as RNA-protein interaction modeling.

Next steps

Moving forward, the team envisions extending their AI strategy to encompass protein-RNA interactions, an area where reliable AI approaches for high-quality protein-RNA complex structure prediction are currently absent. Such tools are highly relevant for RNA function annotation and RNA drug discovery.

In addition, the team hopes to further improve DRfold's accuracy in single-chain RNA structure prediction. One of the inherent barriers stems from the limited availability of experimental RNA structures, which impacts the accuracy of the deep learning models, especially for large-sized RNAs (approximately more than 200 nucleotides).

Novel strategies and ideas are needed to break through the bottleneck of high-accuracy RNA structure predictions, and the researchers are currently working on it with encouraging progress.

More information: Yang Li et al, Integrating end-to-end learning with deep geometrical potentials for ab initio RNA structure prediction, *Nature Communications* (2023). [DOI: 10.1038/s41467-023-41303-9](https://doi.org/10.1038/s41467-023-41303-9)

Provided by National University of Singapore

Citation: Novel AI-based approach for more accurate RNA 3D structure prediction (2023, December 21) retrieved 27 April 2024 from <https://phys.org/news/2023-12-ai-based-approach-accurate-rna-3d.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.