Molecular Biology Laboratory

Bioinformatics and Genomics Lab.

3. Structure Prediction of RNA & Proteins

TA
Junseob Han, Hyunseok Song

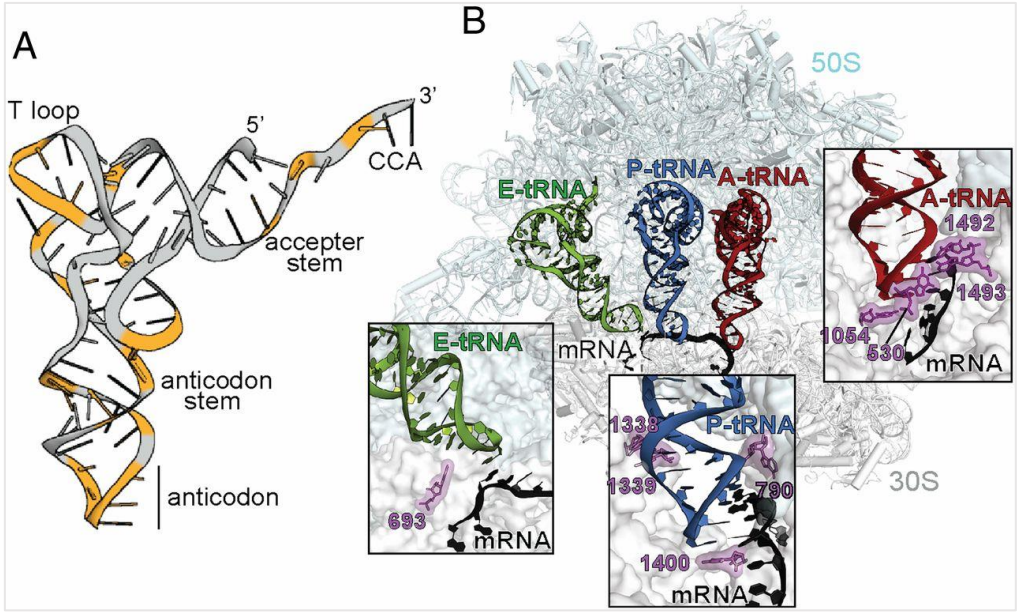
Contact
Junseob Han
010.2113.6458
hljs502@gmail.com

Goal of This Week

1. To know why structure prediction of RNA and protein is important
2. To know how to predict RNA structure (mfold, RNAfold)
3. To know how to predict protein structure (PSI-PRED, Phyre2, AlphaFold)

RNA Structure

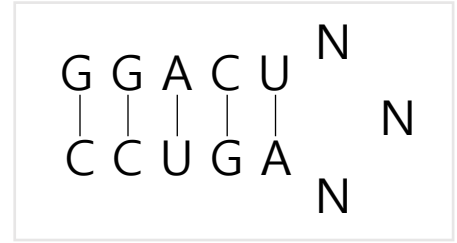
- RNA is known to function as the intermediate molecule of DNA and protein, but some non-coding RNAs act themselves
  - Transfer RNA (tRNA), Ribosomal RNA (rRNA), long non-coding RNA (lncRNA) etc.
- These RNAs have a unique structure to act unique function

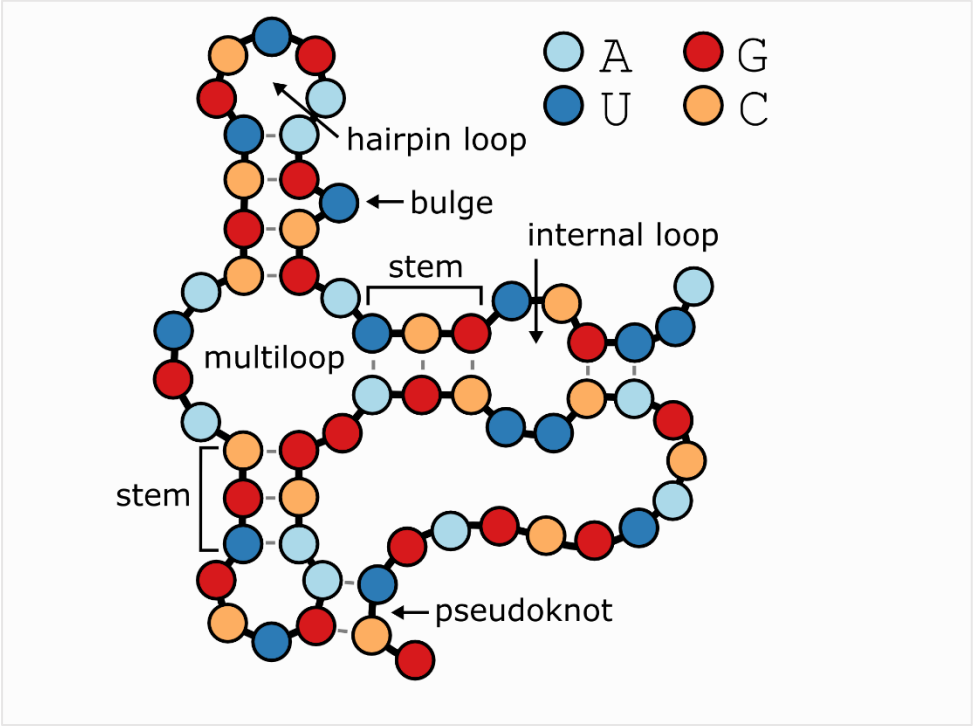


Samuel Hong et al., PNAS, 2018

RNA Secondary Structure

- "Palindrome" is a word that reads the same backward as forwarding
- Some RNAs have local palindrome sequence, so they make a unique structure
  - GGACUNNNAGUCC: This sequence makes the hairpin loop
- There are many types of secondary structures in RNA
  - Hairpin loop, bulge, stem, pseudoknot, etc.



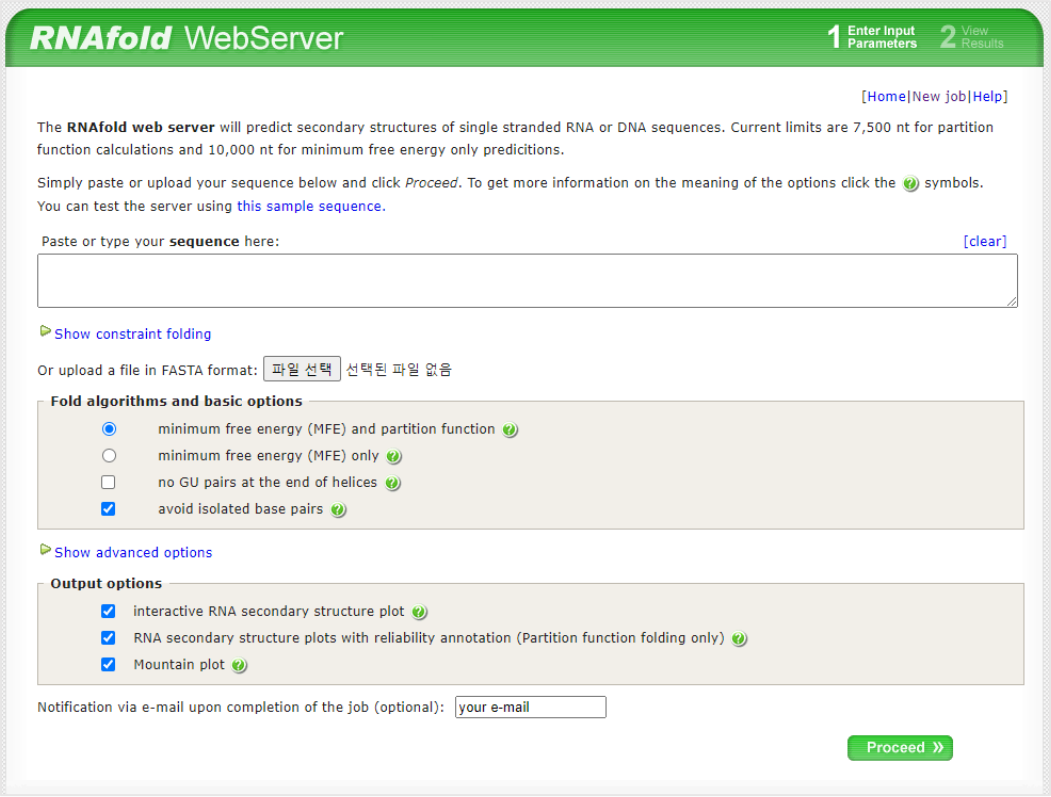


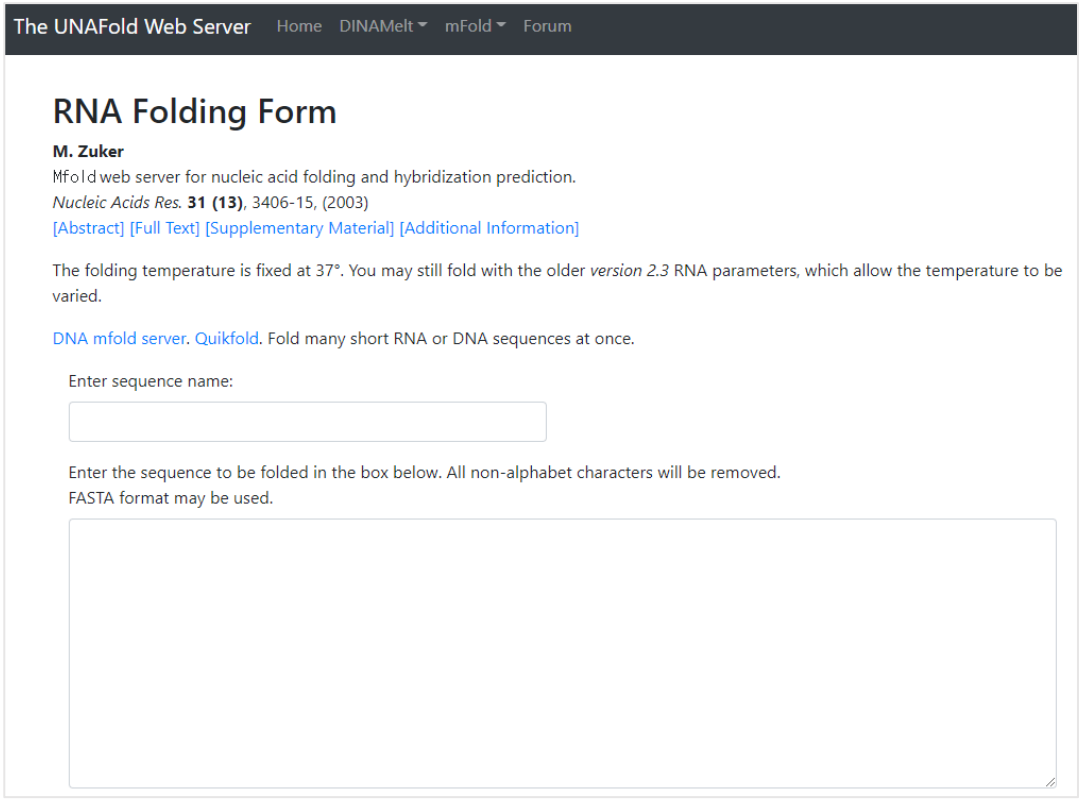
RNA Structure Prediction

- Base pair maximization
  - Find all possible combinations of structures and select the one with the most base pairs
  - Many base pairs and stable structures are separate issues
  - In this algorithm, it is hard to find a suitable structure
- Minimum free energy finding
  - Molecules in cells also follow thermodynamic laws, so suitable RNA structures might have low free energy than other structures
  - The free energy of each base pair is calculated by experimental method and calculate the free energy of the predicted structure, then select the RNA structure which has minimum free energy

RNA Structure Prediction - RNAfold, mfold

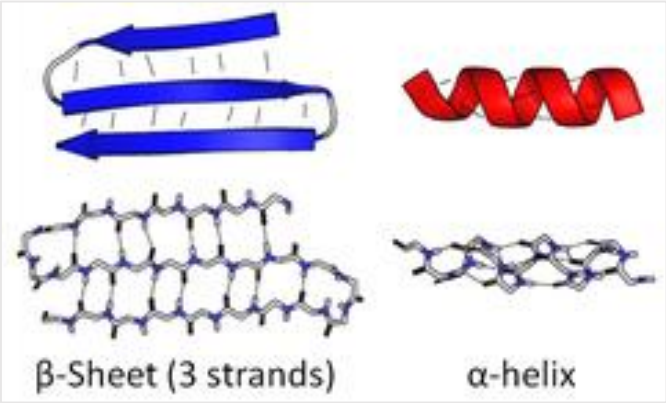
- RNAfold and mfold use minimum free energy finding method
- They use dynamic programming for the minimum free energy of each structure
- They also consider neighborhood sequence and loop structures too





Protein Structure

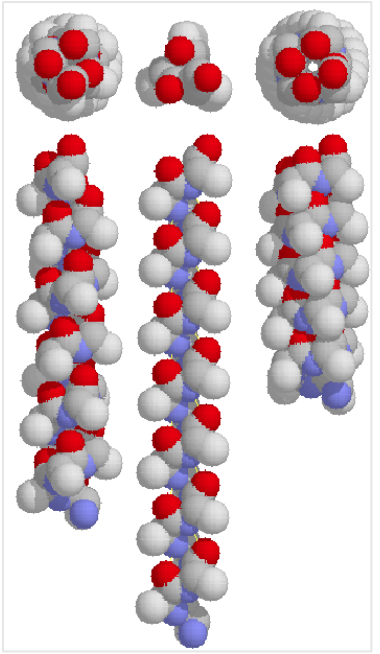
- We can almost exactly predict their functions by structures because they are highly correlated
  - If we predict protein structures exactly, we can know their function and use them for treatment
- Proteins have various amino acids and physical bonds, so it is hard to predict exact structures
  - Protein's secondary structures are largely divided into two categories (α-helix, β-sheet), but in detail, they have more various structures



Protein secondary structure, Wikipedia

DSSP Classification

- 3-turn helix
- 4-turn helix
- 5-turn helix
- β-bridge
- β-sheet
- Bend
- Hydrogen bonded turn



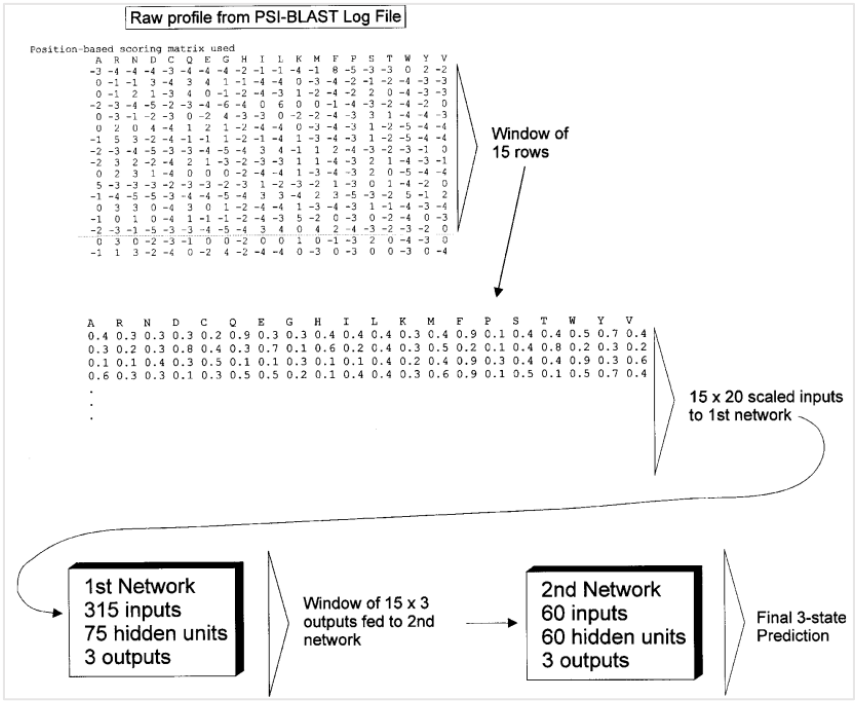
Helices, Principles of Protein Structure

Protein Structure Prediction

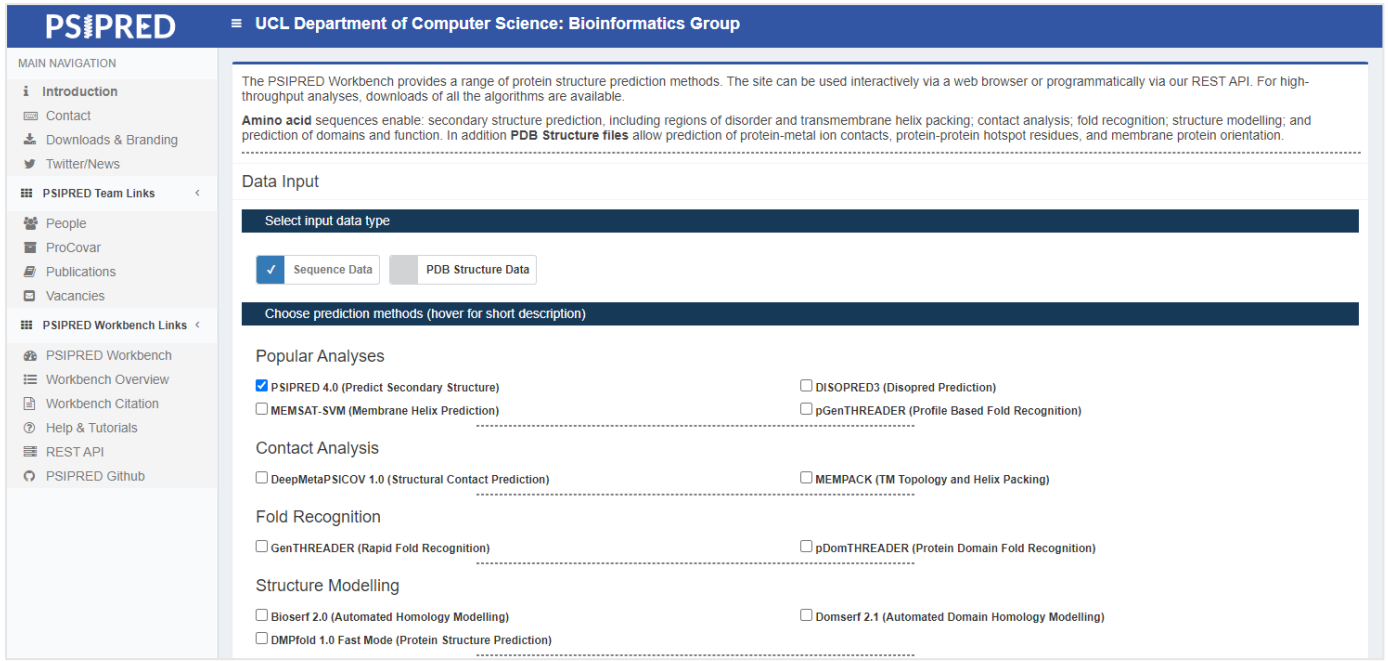
- Chou-Fasman algorithm
  - Predict structures by calculating the probability that amino acids form α-helix or β-sheet based on known structures
  - This algorithm doesn't consider neighborhood amino acids, so accuracy is low (Approximately 50%)
- Garnier-Osquthorpe-Robson (GOR) algorithm
  - In addition to the Chou-Fasman algorithm, predict structures by considering neighborhood amino acids
  - This algorithm shows a little better accuracy (Approximately 65%)
- Neural network machine learning algorithm
  - Based on neural network machine learning, find structures that have a similar amino acid sequence
  - "PSI-PRED" uses this algorithm and shows much better accuracy than other algorithms (Approximately 77%)

Protein Structure Prediction - PSI-PRED

- "PSI-PRED" perform multiple sequence alignment with "PSI-BLAST", and calculates position-specific scoring matrix (PSSM)
- It builds 20 X 15 size matrix (15 amino acids bin) and predicts structure by using the pre-built neural network model

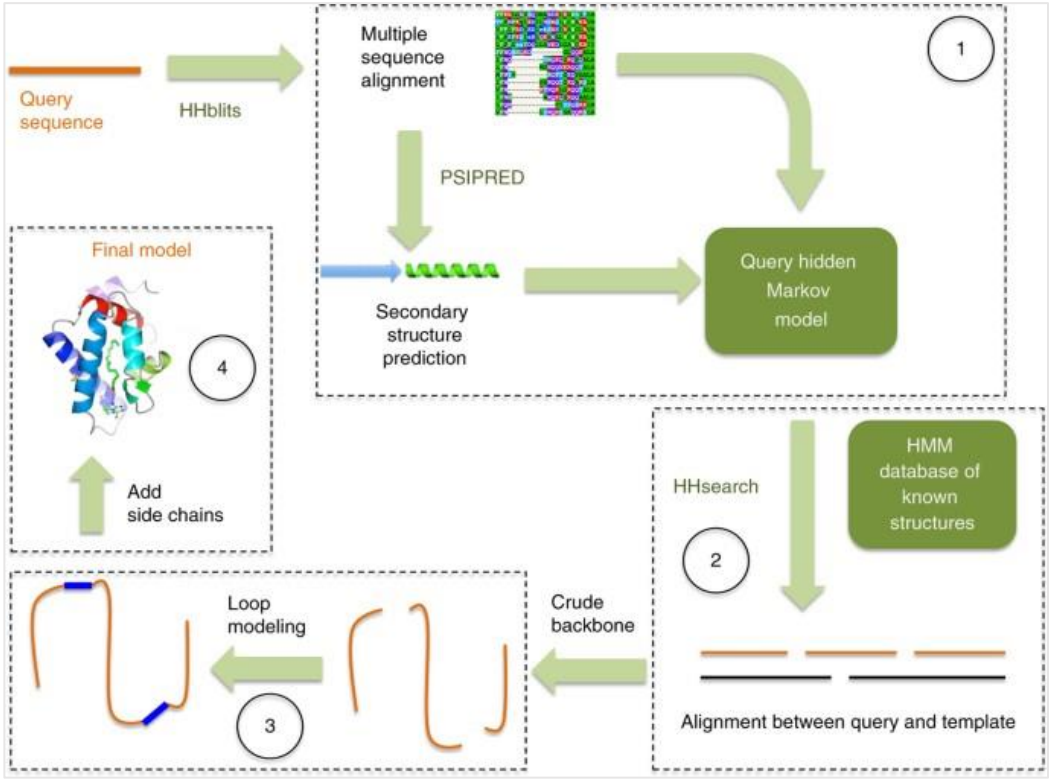


David T. Jones, Journal of molecular biology, 1999

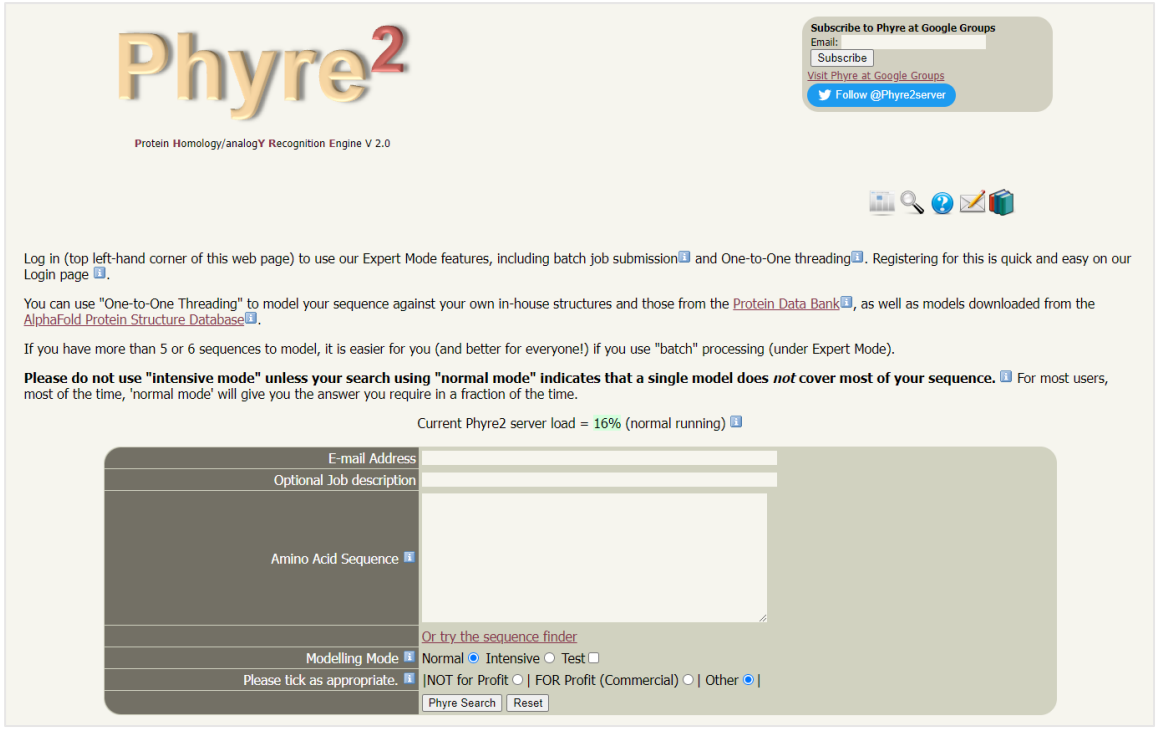


Protein 3D Structure Prediction - Phyre2

- "Phyre2" is used to predict the 3D structure of proteins
- "Phyre2" uses "PSI-PRED" for predicting secondary structures and a large database of 3D structures for finding similar sequences with our query sequence



Lawrence A Kelley et al., Nature Protocols, 2015



Further View - Structure Prediction with Deep Learning

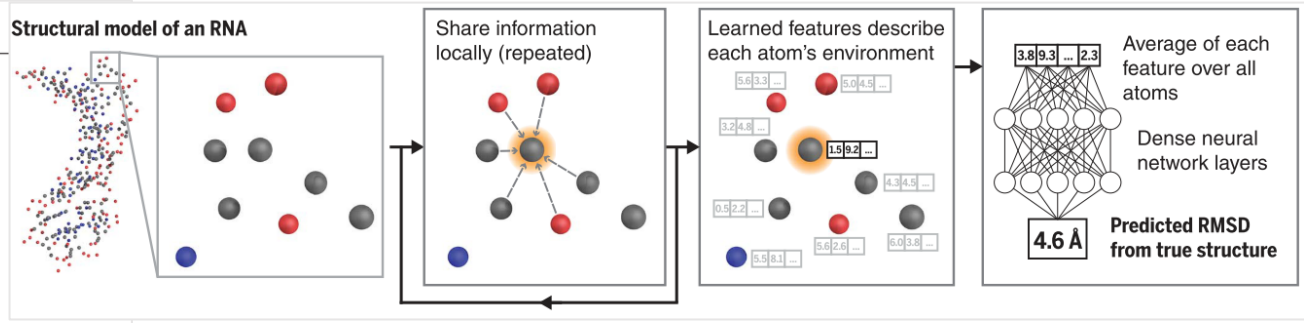
RESEARCH

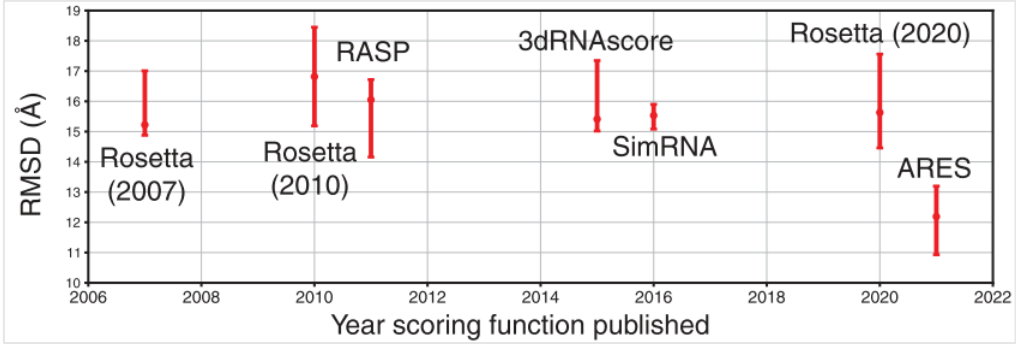
RNA

Geometric deep learning of RNA structure

Raphael J. L. Townshend[1]†‡, Stephan Eismann[1,2]†, Andrew M. Watkins[3]†, Ramya Rangan[3,4], Maria Karelina[1,4], Rhiju Das[3,5]*, Ron O. Dror[1,6,7,8]*

RNA molecules adopt three-dimensional structures that are critical to their function and of interest in drug discovery. Few RNA structures are known, however, and predicting them computationally has proven challenging. We introduce a machine learning approach that enables identification of accurate structural models without assumptions about their defining characteristics, despite being trained with only 18 known RNA structures. The resulting scoring function, the Atomic Rotationally Equivariant Scorer (ARES), substantially outperforms previous methods and consistently produces the best results in community-wide blind RNA structure prediction challenges. By learning effectively even from a small amount of data, our approach overcomes a major limitation of standard deep neural networks. Because it uses only atomic coordinates as inputs and incorporates no RNA-specific information, this approach is applicable to diverse problems in structural biology, chemistry, materials science, and beyond.





Further View - Structure Prediction with Deep Learning

Article

Highly accurate protein structure prediction with AlphaFold

https://doi.org/10.1038/s41586-021-03819-2

Received: 11 May 2021

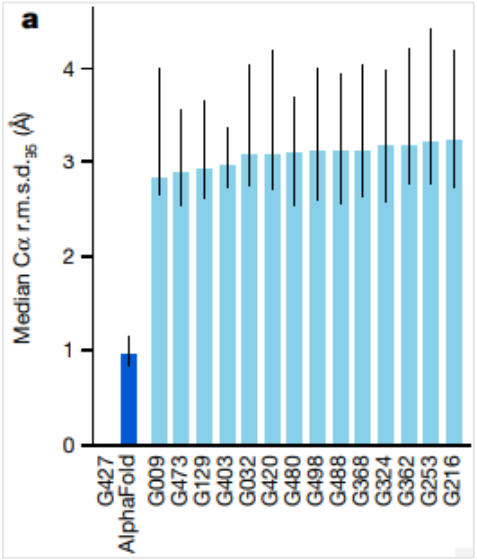
Accepted: 12 July 2021

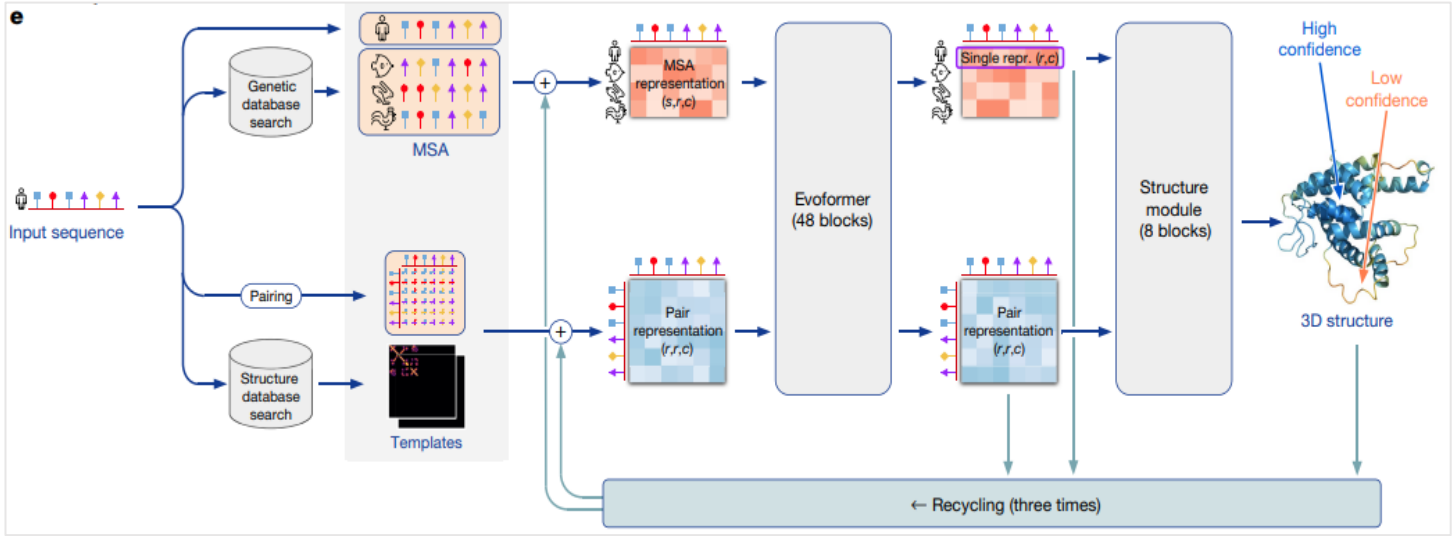
Published online: 15 July 2021

Open access

Check for updates

John Jumper[1,4]✉, Richard Evans[1,4], Alexander Pritzel[1,4], Tim Green[1,4], Michael Figurnov[1,4], Olaf Ronneberger[1,4], Kathryn Tunyasuvunakool[1,4], Russ Bates[1,4], Augustin Žídek[1,4], Anna Potapenko[1,4], Alex Bridgland[1,4], Clemens Meyer[1,4], Simon A. A. Kohl[1,4], Andrew J. Ballard[1,4], Andrew Cowie[1,4], Bernardino Romera-Paredes[1,4], Stanislav Nikolov[1,4], Rishub Jain[1,4], Jonas Adler[1], Trevor Back[1], Stig Petersen[1], David Reiman[1], Ellen Clancy[1], Michal Zielinski[1], Martin Steinegger[2,3], Michalina Pacholska[1], Tamas Berghammer[1], Sebastian Bodenstein[1], David Silver[1], Oriol Vinyals[1], Andrew W. Senior[1], Koray Kavukcuoglu[1], Pushmeet Kohli[1] & Demis Hassabis[1,4]✉





Practical Exercise

1. Practice how to predict RNA structure
    - Predict the structure of "tRNA Val" with "RNAfold" and "mfold"
2. Practice how to predict protein structures
    - Predict the structure of green fluorescence protein (GFP) with "PSI-PRED"
    - Predict the 3D structure of GFP with “ColabFold" and compare with X-ray crystallography of GFP