

# 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

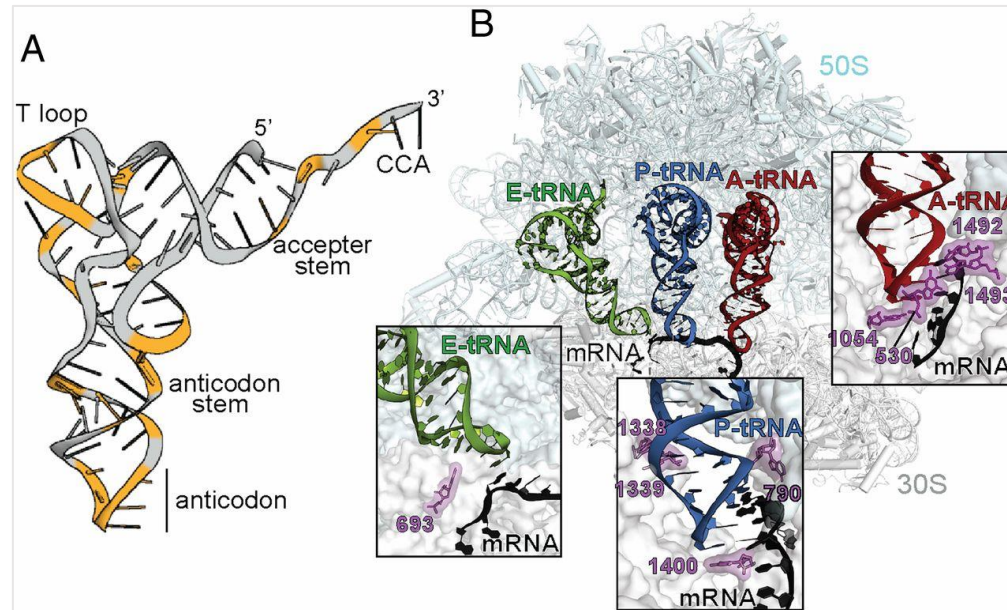
# Goal of This Week

---

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

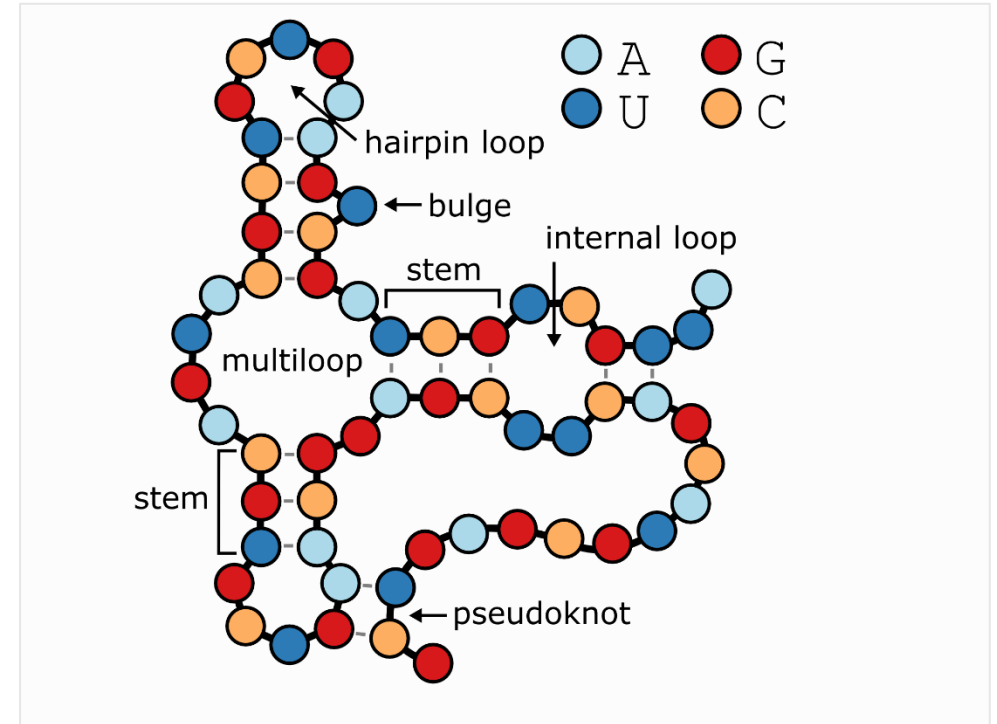
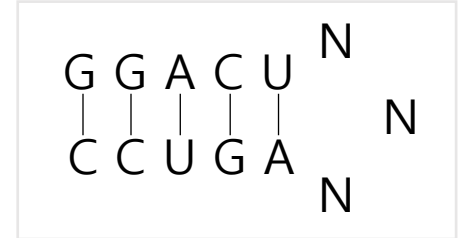
- RNA is known to function as the intermediate molecule of DNA and protein, but some non-coding RNAs act themselves
  - Transfer RNA (tRNA), Ribosomal RNA (rRNA), long non-coding RNA (lncRNA) etc.
- These RNAs have a unique structure to act unique function



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

**RNAfold WebServer** 1 Enter Input Parameters 2 View Results

[\[Home\]](#) [\[New job\]](#) [\[Help\]](#)

The **RNAfold web server** will predict secondary structures of single stranded RNA or DNA sequences. Current limits are 7,500 nt for partition function calculations and 10,000 nt for minimum free energy only predictions.

Simply paste or upload your sequence below and click *Proceed*. To get more information on the meaning of the options click the symbols. You can test the server using [this sample sequence](#).

Paste or type your **sequence** here: [\[clear\]](#)

[▶ Show constraint folding](#)

Or upload a file in FASTA format:  선택된 파일 없음

**Fold algorithms and basic options**

- minimum free energy (MFE) and partition function
- minimum free energy (MFE) only
- no GU pairs at the end of helices
- avoid isolated base pairs

[▶ Show advanced options](#)

**Output options**

- interactive RNA secondary structure plot
- RNA secondary structure plots with reliability annotation (Partition function folding only)
- Mountain plot

Notification via e-mail upon completion of the job (optional):

The UNAFold Web Server [Home](#) [DINAMelt](#) [mFold](#) [Forum](#)

## RNA Folding Form

**M. Zuker**  
Mfold web server for nucleic acid folding and hybridization prediction.  
*Nucleic Acids Res.* **31 (13)**, 3406-15, (2003)  
[\[Abstract\]](#) [\[Full Text\]](#) [\[Supplementary Material\]](#) [\[Additional Information\]](#)

The folding temperature is fixed at 37°. You may still fold with the older *version 2.3* RNA parameters, which allow the temperature to be varied.

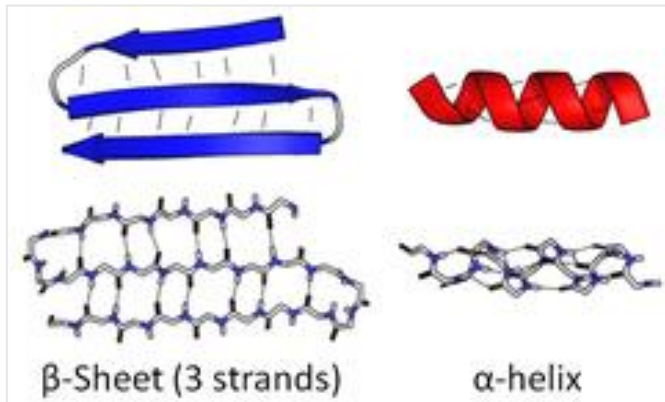
[DNA mfold server](#). [Quikfold](#). Fold many short RNA or DNA sequences at once.

Enter sequence name:

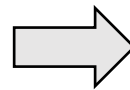
Enter the sequence to be folded in the box below. All non-alphabet characters will be removed.  
FASTA format may be used.

# 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 ( $\alpha$ -helix,  $\beta$ -sheet), but in detail, they have more various structures

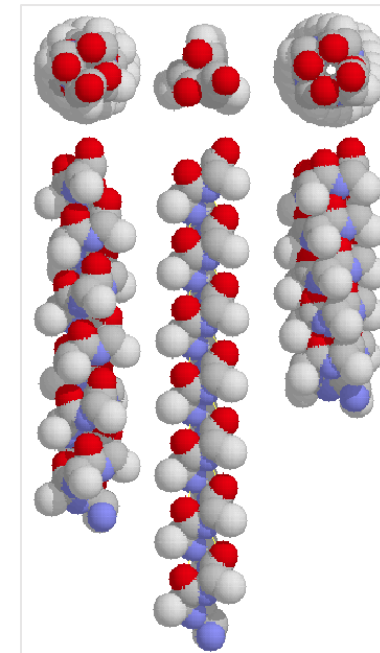


Protein secondary structure, Wikipedia



## DSSP Classification

- 3-turn helix
- 4-turn helix
- 5-turn helix
- $\beta$ -bridge
- $\beta$ -sheet
- Bend
- Hydrogen bonded turn



Helices, Principles of Protein Structure

# Protein Structure Prediction

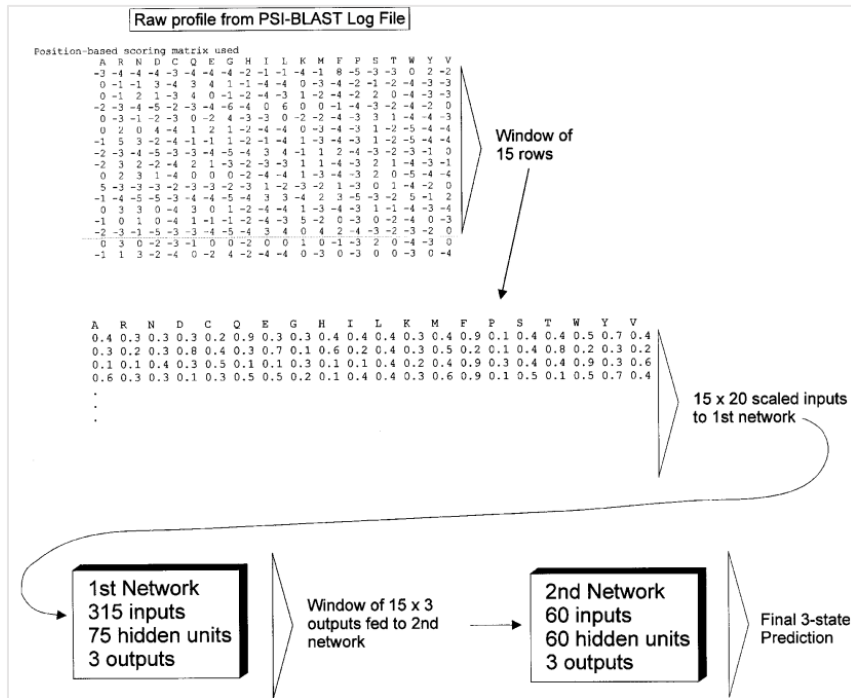
---

- Chou-Fasman algorithm
  - Predict structures by calculating the probability that amino acids form  $\alpha$ -helix or  $\beta$ -sheet based on known structures
  - This algorithm doesn't consider neighborhood amino acids, so accuracy is low (Approximately 50%)
- Garnier-Osquithorpe-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

PSIPRED UCL Department of Computer Science: Bioinformatics Group

MAIN NAVIGATION

- Introduction
- Contact
- Downloads & Branding
- Twitter/News
- PSIPRED Team Links
- People
- ProCovar
- Publications
- Vacancies
- PSIPRED Workbench Links
- PSIPRED Workbench
- Workbench Overview
- Workbench Citation
- Help & Tutorials
- REST API
- PSIPRED Github

The PSIPRED Workbench provides a range of protein structure prediction methods. The site can be used interactively via a web browser or programmatically via our REST API. For high-throughput analyses, downloads of all the algorithms are available.

Amino acid sequences enable: secondary structure prediction, including regions of disorder and transmembrane helix packing; contact analysis; fold recognition; structure modelling; and prediction of domains and function. In addition PDB Structure files allow prediction of protein-metal ion contacts, protein-protein hotspot residues, and membrane protein orientation.

Data Input

Select input data type

Sequence Data  PDB Structure Data

Choose prediction methods (hover for short description)

Popular Analyses

- PSIPRED 4.0 (Predict Secondary Structure)
- MEMSAT-SVM (Membrane Helix Prediction)
- DISOPRED3 (Disopred Prediction)
- pGenTHREADER (Profile Based Fold Recognition)

Contact Analysis

- DeepMetaPSICOV 1.0 (Structural Contact Prediction)
- MEMPACK (TM Topology and Helix Packing)

Fold Recognition

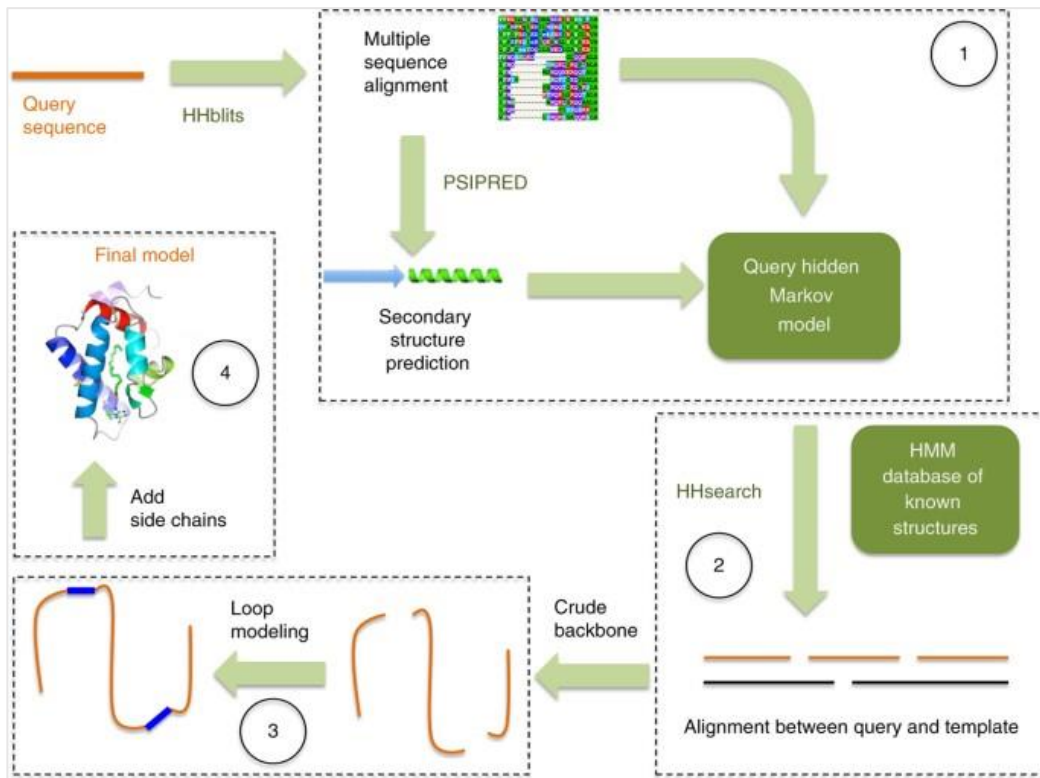
- GenTHREADER (Rapid Fold Recognition)
- pDomTHREADER (Protein Domain Fold Recognition)

Structure Modelling

- Bioserf 2.0 (Automated Homology Modelling)
- Domserf 2.1 (Automated Domain Homology Modelling)
- DMPfold 1.0 Fast Mode (Protein Structure Prediction)

# 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

The screenshot shows the Phyre2 web interface. At the top, the logo "Phyre<sup>2</sup>" is displayed with the subtitle "Protein Homology/analogy Recognition Engine V 2.0". There are links to "Subscribe to Phyre at Google Groups", "Visit Phyre at Google Groups", and "Follow @Phyre2server". Below the logo, there are navigation icons for home, search, help, and email. The main content area contains text about logging in for Expert Mode features and using "One-to-One Threading". A warning states: "Please do not use 'intensive mode' unless your search using 'normal mode' indicates that a single model does not cover most of your sequence." The current server load is shown as 16%. The bottom section is a form for submitting a query, with fields for "E-mail Address", "Optional Job description", and "Amino Acid Sequence". There are also options for "Modelling Mode" (Normal, Intensive, Test) and "Please tick as appropriate" (NOT for Profit, FOR Profit (Commercial), Other). Buttons for "Phyre Search" and "Reset" are at the bottom.

# Further View - Structure Prediction with Deep Learning

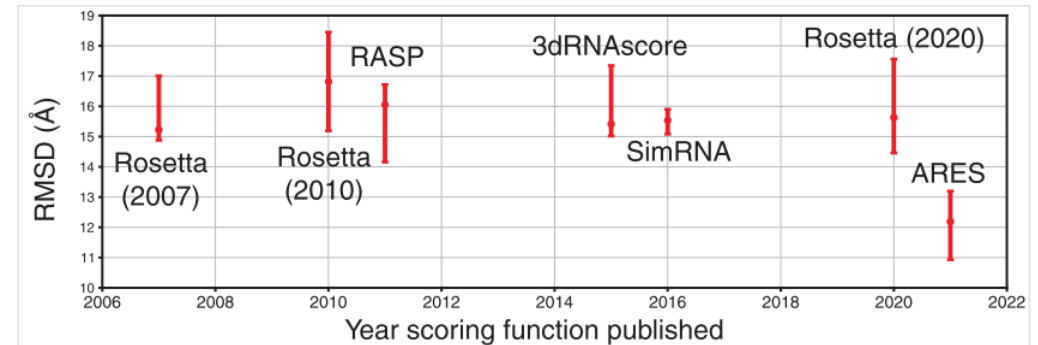
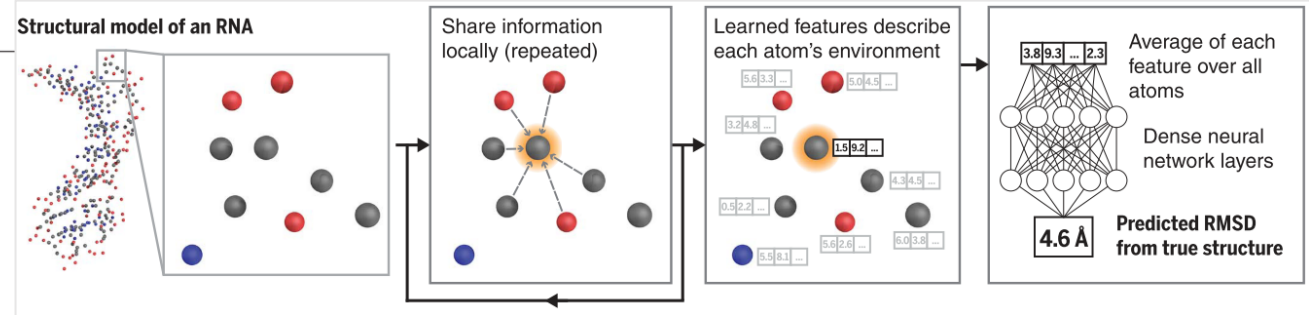
## RESEARCH

### RNA

## Geometric deep learning of RNA structure

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

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

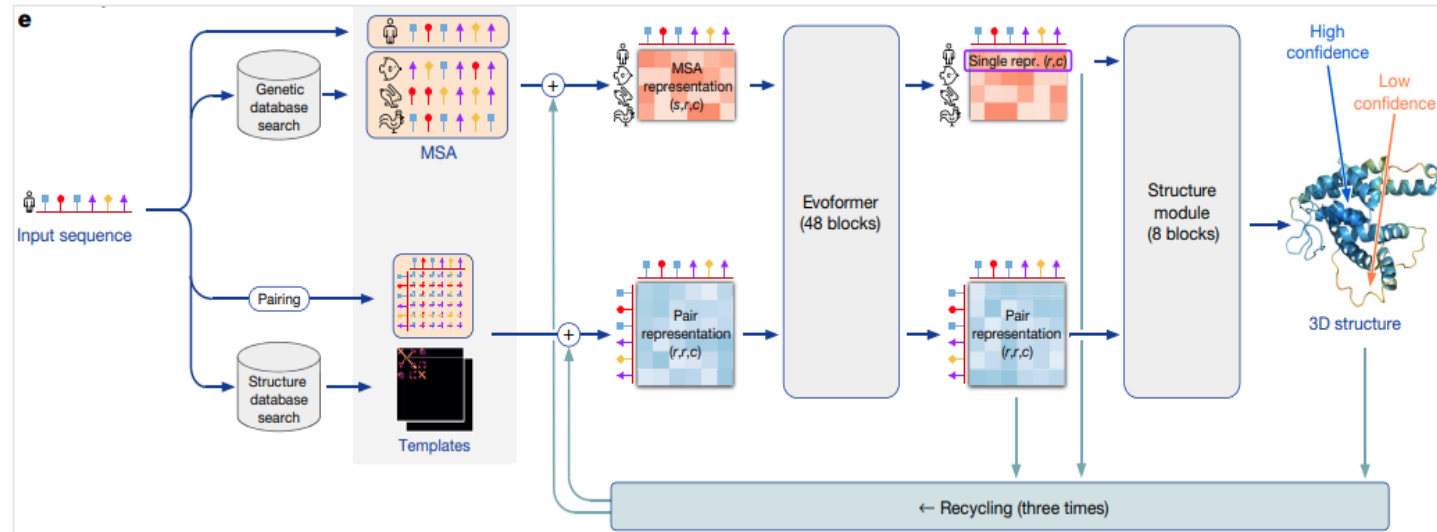
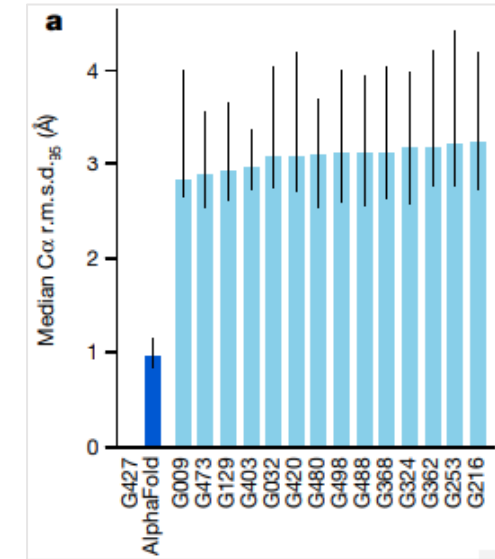
Accepted: 12 July 2021

Published online: 15 July 2021

Open access

 Check for updates

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



# 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