

Molecular Biology Laboratory

Bioinformatics and Genomics Lab.

Week3. Structure Prediction of RNA & Proteins

1. Practice how to predict RNA structure

- Predict the structure of "tRNA Val" with "RNAfold" and "mfold"
 - Search "NCBI" in google and access NCBI.

A screenshot of a Google search for "ncbi". The search bar contains "ncbi" and the search button is visible. Below the search bar, there are navigation options: "전체" (All), "이미지" (Images), "도서" (Books), "동영상" (Videos), "쇼핑" (Shopping), and "더보기" (More). The search results show approximately 608,000,000 results in 0.52 seconds. The top result is "National Center for Biotechnology Information" with a red arrow pointing to the link. Below the link, there is a brief description: "Welcome to NCBI. The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information." There are also links for "BLAST", "PubMed", "Nucleotide", and "Gene". On the right side, there is a sidebar with the NCBI logo and Korean text: "미국 국립생물공학정보센터 회사" and "미국 국립생물공학정보센터는 미국 보건성 산하 국립 의학도서관의 운영 분야 중 하나이다. NCBI는 1988년 미국 메릴랜드주에 설립되었다. 위키백과" (The National Center for Biotechnology Information is one of the operating fields of the National Library of Medicine, a part of the U.S. Department of Health and Human Services. NCBI was established in Maryland, USA in 1988. Wikipedia).

- Search "human tRNA-Val" and click "Gene".

A screenshot of the NCBI search results for "human tRNA-Val". The search bar at the top contains "human tRNA-Val" and the "Search" button is visible. Below the search bar, it says "Results found in 15 databases". The main result is for the gene "TRNV - tRNA-Val" from "Homo sapiens (human)". It also lists "Also known as: MTTV" and "Gene ID: 4577". There is a "PubMed (7)" link and a "Genome Data Viewer" button. Below the main result, there are three columns of data: "Literature", "Genes", and "Proteins". The "Genes" column has a red arrow pointing to the "Gene" entry. The "Proteins" column has a "747" next to the "Protein" entry.

Literature	Genes	Proteins
Bookshelf	Gene	Conserved Domains
MeSH	GEO DataSets	Identical Protein Groups
NLM Catalog	GEO Profiles	Protein
PubMed	HomoloGene	Protein Family Models
PubMed Central	PopSet	Structure
33	214	0
0	0	0
0	0	747
191	0	0
13,754	464	8

- Select one from the list. We will use 3rd one, "tRNA-Val (anticodon AAC) 1-4" in the practical exercise.

Search results

Items: 1 to 20 of 214 << First < Prev Page 1 of 11 Next > Last >>

[See also 166 discontinued or replaced items.](#)

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> TRV-CAC1-7 ID: 100189228	tRNA-Val (anticodon CAC) 1-7 [<i>Homo sapiens</i> (human)]	Chromosome 1, NC_000001.11 (149712552..149712624, complement)	TRNAV22, TRV-CAC5-1	
<input type="checkbox"/> MT-TV ID: 4577	mitochondrially encoded tRNA valine [<i>Homo sapiens</i> (human)]	Chromosome MT, NC_012920.1 (1602..1670)	MTTV, TRNV	
<input type="checkbox"/> TRV-AAC1-4 ID: 7239	tRNA-Val (anticodon AAC) 1-4 [<i>Homo sapiens</i> (human)]	Chromosome 5, NC_000005.10 (181218270..181218342, complement)	TRNAV1, TRV, TRV1	189921
<input type="checkbox"/> TRV-CAC3-1 ID: 100189416	tRNA-Val (anticodon CAC) 3-1 [<i>Homo sapiens</i> (human)]	Chromosome 19, NC_000019.10 (4724635..4724707, complement)	TRNAV32	615304
<input type="checkbox"/> TRV-CAC1-6 ID: 100189227	tRNA-Val (anticodon CAC) 1-6 [<i>Homo sapiens</i> (human)]	Chromosome 6, NC_000006.12 (26538054..26538126)	TRNAV21	615306

- Click "FASTA" to get the gene sequence of tRNA-Val.

Homo sapiens chromosome 5, GRCh38.p14 Primary Assembly

NCBI Reference Sequence: NC_000005.10

[GenBank](#) [Graphics](#)

```
>NC_000005.10:c181218342-181218270 Homo sapiens chromosome 5, GRCh38.p14 Primary Assembly
GTTCCGTAGTGTAGTGGTTATCACGTTCCGCTAACACGCGAAAGGTCCCGGTCGAAACCGGGCGGAA
ACA
```

- Search "RNAfold" in google and access RNAfold.

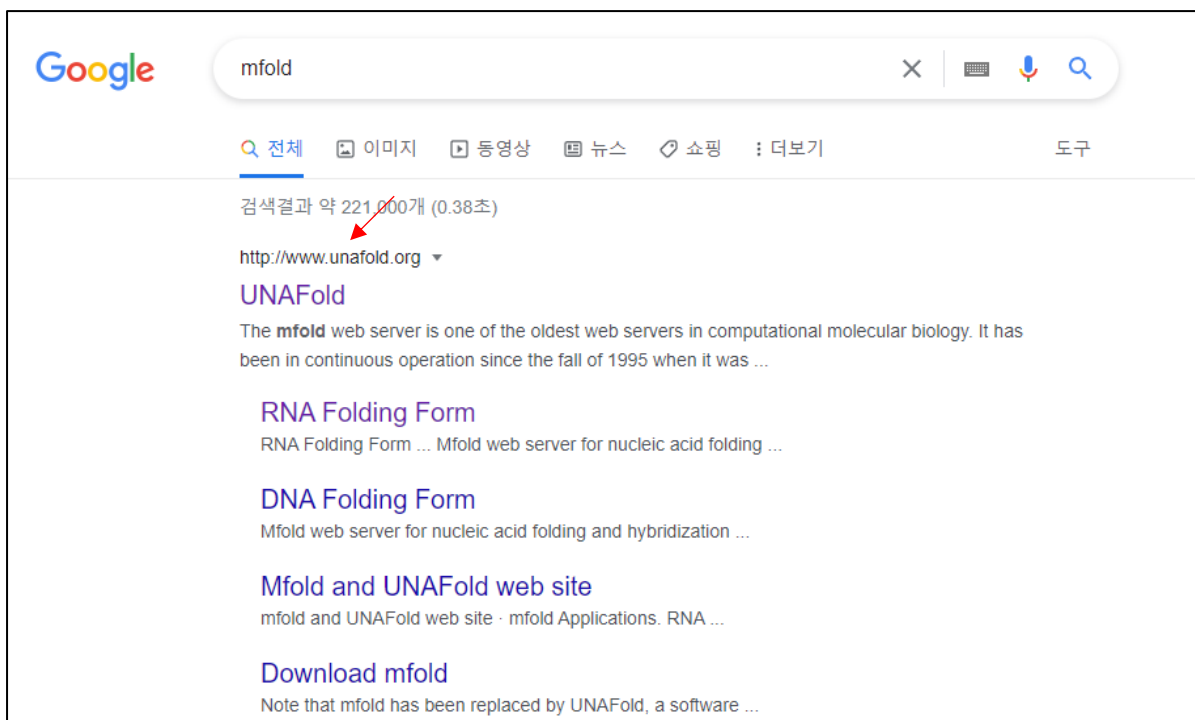
Google search for "rnafold".

Search results: 약 89,600개 (0.49초)

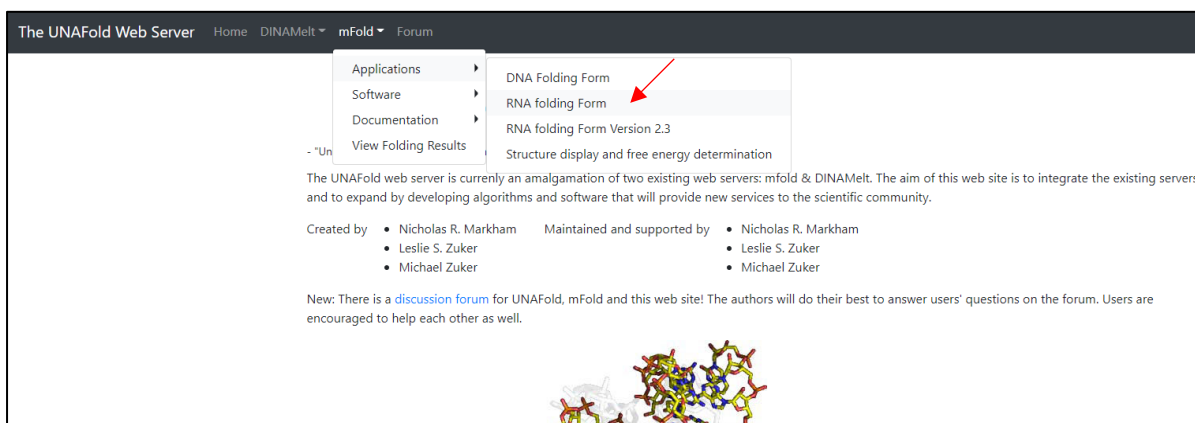
http://rna.tbi.univie.ac.at > cgi-bin / RNAWebSuite > R...
RNAfold web server
 The **RNAfold** web server will predict secondary structures of single stranded RNA or DNA sequences. Current limits are 7,500 nt for partition function ...

http://rna.tbi.univie.ac.at >
ViennaRNA Web Services
 Thermodynamic Structure Prediction. **RNAfold** Server ...predicts minimum free energy structures and base pair probabilities from single RNA or DNA sequences.

- Search "mfold" in google and access UNAFold



- Hover the mouse on the "mFold" - "Applications" and click "RNA folding Form"



- Copy and paste the tRNA sequence to the box, go to below and click "Fold RNA" for RNA structure prediction.

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.

```
>NC_000005.10:c181218342-181218270 Homo sapiens chromosome 5, GRCh38.p14 Primary Assembly
GTTTCGGTAGTGTAGTGGTTATCAAGTTCGCCTAACACGCGAAAGGTCGCCGGTTCGAAACCGGGCGGAA
ACA
```

[Format Sequence](#) [Clear Constraints](#) [Check Constraints](#)

- In the result, we can see predicted structures and their free energy. Click "pdf" of "Structure 1" then we can see the structure which has minimum free energy

Output

The *energy dot plot* for NC_000005 10:c181218342-181218270 Homo sapiens chromosome 5, GRCh38 p1. ([Definition](#))
File formats: [Text](#), [PostScript](#), [pdf](#), [png](#), [jpg](#)
Computed Structures: ([File Formats](#))
The computed foldings contain 32 base pairs out of 44 (72.7%) in the *energy dot plot*.

Extra files: [sorted ct file](#); [h-num](#) values; [p-num](#) values; [log file](#) for main computations.

Download all foldings

Choose *zipped* file: or *gzipped tar* file: [PostScript](#)

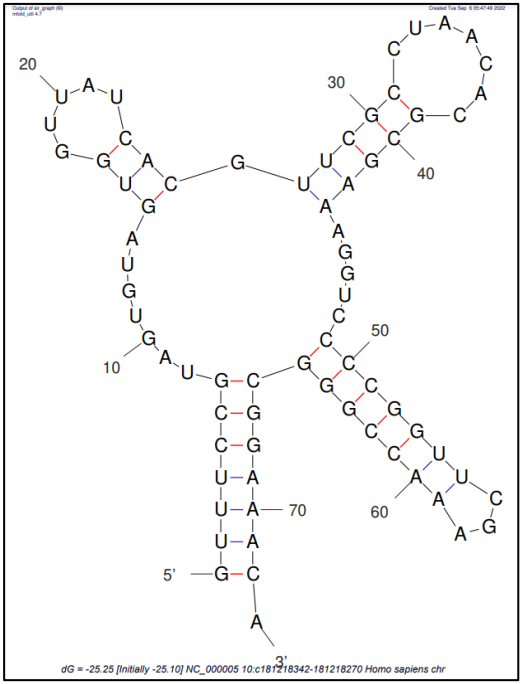
View ss-count information

([Definition](#)) ([ss-count file](#)) ss value = 1.22 ± 1.23
Averaging window Magnification Base to magnify about Plot format [PostScript](#)

View Individual Structures:

[Circular structure Plots](#)

- ◆ **Structure 1** : Initial $\Delta G = -25.10$ kcal/mol, ([Thermodynamic Details](#)).
Different file formats: [PostScript](#), [pdf](#), [png](#), [jpg](#), [.ct file](#), [Vienna](#), [RNAML](#), [RnaViz ct](#), [Mac ct](#), [RNA draw](#), [XRNA ss](#).
- ◆ **Structure 2** : Initial $\Delta G = -24.80$ kcal/mol, ([Thermodynamic Details](#)).
Different file formats: [PostScript](#), [pdf](#), [png](#), [jpg](#), [.ct file](#), [Vienna](#), [RNAML](#), [RnaViz ct](#), [Mac ct](#), [RNA draw](#), [XRNA ss](#).
- ◆ **Structure 3** : Initial $\Delta G = -24.50$ kcal/mol, ([Thermodynamic Details](#)).
Different file formats: [PostScript](#), [pdf](#), [png](#), [jpg](#), [.ct file](#), [Vienna](#), [RNAML](#), [RnaViz ct](#), [Mac ct](#), [RNA draw](#), [XRNA ss](#).



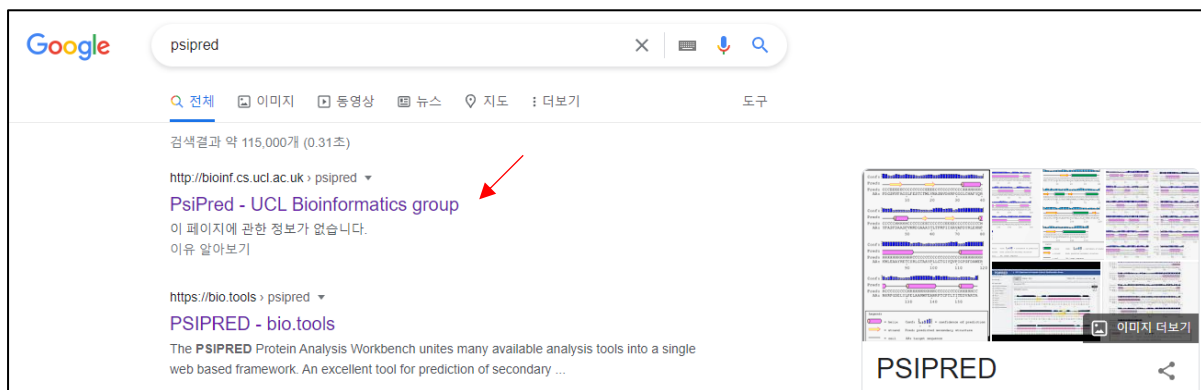
2. Practice how to predict proteins structure

- Predict the structure of green fluorescence protein (GFP) with "PSI-PRED"

- Go to UniProt and get the protein sequence of GFP

```
>sp|P42212|GFP_AEQVI Green fluorescent protein OS=Aequorea victoria OX=6100 GN=GFP PE=1 SV=1
MSKGEELFTGVVPILVELDGDVNGHKFVSVSGEGEGDATYGKLT LKFICTTGKLPVWPPTL
VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV
NRIELKGI DFKE DGNILGHKLEYNYN SHNVYIMADKQKNGIKVNFKIRHNI EDGSVQLAD
HYQQNTPI GDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITHGMDELYK
```

- Search "psipred" in google and access PSI-PRED.



- Copy and paste the GFP sequence to the box, go to below and click "Submit" for protein secondary structure prediction.

Submission details

Protein Sequence

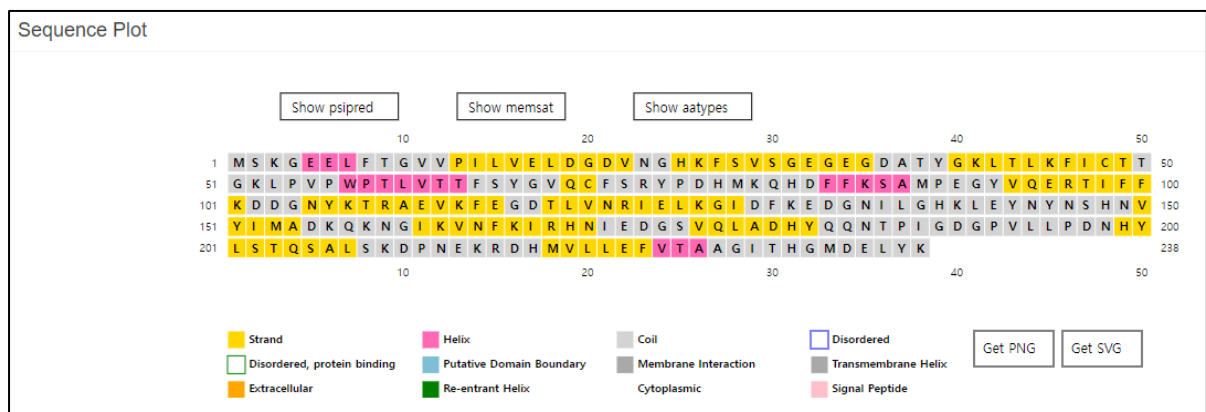
>sp|P42212|GFP_AEQVI Green fluorescent protein OS=Aequorea victoria OX=6100 GN=GFP PE=1 SV=1
 MSKGEELFTGVVPILVELDGDVNGHKFVSVSGEGEGDATYGKLT LKFICTTGKLPVWPPTL
 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV

Help...
 If you wish to test these services follow this link to retrieve a test fasta sequence.

Job name

Email (optional)

- In the result, we can separate α -helix, β -sheet, and coil region.



- Predict the 3D structure of GFP and compare with X-ray crystallography image

➤ Search “phyre2” in google and access Phyre2

(In the current situation, google directs an error page. You can access by this hyperlink

<http://www.sbg.bio.ic.ac.uk/phyre2>)

➤ Write your e-mail for receiving the result, and job name, and copy and paste the GFP sequence to the box without a header of fasta format. And then click “Phyre Search”.

- Wait until to finish their prediction.


Phyre²

Job Status

Email	hljs502@gmail.com
Job Description	GFP_____
Unique Job ID	ccf86e404f8972d4
Date	Thu Nov 16 06:37:30 GMT 2023

Estimated total processing time: 1.5 hours ± 1.4 hours [i](#)

1. Finding homologues with PSI-Blast



A link to results will be mailed to you when the job is finished
Or bookmark this page to return to it at any time

- In the result, we can see the 3D structure of GFP, and some of the data

Phyre²

Email	hljs502@gmail.com
Description	GFP_____
Date	Thu Nov 16 06:37:30 GMT 2023
Unique Job ID	ccf86e404f8972d4
Sequence	MSKGEEELFTG ... Download FASTA
Job type	normal
Job Expiry	30 days

[Download zip of all results](#)




Image coloured by rainbow N → C terminus
Model dimensions (Å): X:48.957 Y:55.903 Z:48.782

Top model

Model (left) based on template [c6wvdA](#)

Top template information

PDB header: membrane protein
Chain: A; **PDB Molecule:** green fluorescent protein, protein jagunal homolog 1
PDB title: human jagn1
PDB Entry: [PDBe](#) [RCSB](#) [PDBj](#)

Confidence and coverage

Confidence: 100.0% Coverage: 96%

228 residues (96% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

[3D viewing](#)
[Interactive 3D view in JSmol](#)
For other options to view your downloaded structure offline see the [FAQ](#)

- Search "colabfold" in google and access ColabFold

Google search results for "colabfold". The search bar shows "colabfold" and the results page displays approximately 9,500 results. The top result is "AlphaFold2.ipynb - Colaboratory - Google Colab" with a red arrow pointing to the title. Below it is a result from Nature: "ColabFold: making protein folding accessible to all - Nature".

- Copy and paste the GFP sequence to the "query_sequence", set job name, and click "Runtime" - "Run all" for protein 3D structure prediction.

The screenshot shows the Google Colab interface for the "AlphaFold2.ipynb" notebook. The notebook title is "ColabFold: AlphaFold2 using MMseqs2". The code cell [2] contains the following configuration:

```

query_sequence: "MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGEDATYGKLTCLKFICTTGKLPVWPPTL VTTFSYGVQCFS "
jobname: "GFP"
use_amber: 
template_mode: none

```

Red arrows point to the notebook title, the "Runtime" button, and the "Run all" button. The code cell also includes instructions on how to use the "query_sequence" field to specify inter-protein chainbreaks for modeling complexes.

- In the result, we can see the 3D structure of GFP.

Display 3D structure

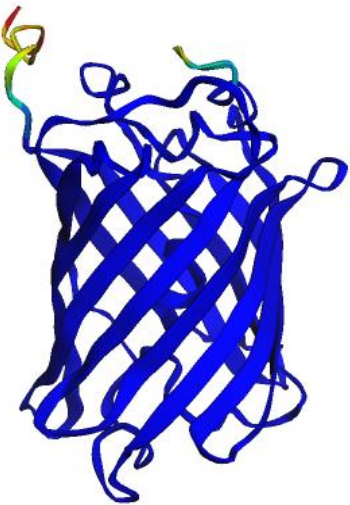
rank_num: 1

color: IDDT

show_sidechains:

show_mainchains:

[코드 표시](#)



pLDDT: ■ Very low (<50) ■ Low (60) ■ OK (70) ■ Confident (80) ■ Very high (>90)

- Compare the predicted structure and X-ray crystal structure of GFP.

