

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 URL "https://www.ncbi.nlm.nih.gov". A red arrow points to this result. Below the main result, there are links for "BLAST", "PubMed", "Nucleotide", and "Gene". On the right side, there is a sidebar with the NCBI logo and text in Korean: "미국 국립생물공학정보센터 회사" and "미국 국립생물공학정보센터는 미국 보건성 산하 국립 의학도서관의 운영 분야 중 하나이다. NCBI는 1988년 미국 메릴랜드주에 설립되었다. 위키백과".

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

A screenshot of the NCBI search results for "human tRNA-Val". The search bar contains "human tRNA-Val" and the search button is visible. The results are displayed in a grid format. The top result is "TRNV - tRNA-Val" from *Homo sapiens (human)*. Below this, there are three columns of results: "Literature", "Genes", and "Proteins". The "Genes" column has a red arrow pointing to the "Gene" result. The "Literature" column shows 33 results in Bookshelf, 0 in MeSH, 0 in NLM Catalog, 191 in PubMed, and 13,754 in PubMed Central. The "Genes" column shows 214 in Gene, 0 in GEO DataSets, 0 in GEO Profiles, 0 in HomoloGene, and 464 in PopSet. The "Proteins" column shows 0 in Conserved Domains, 0 in Identical Protein Groups, 747 in Protein, 0 in Protein Family Models, and 8 in Structure.

- 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.

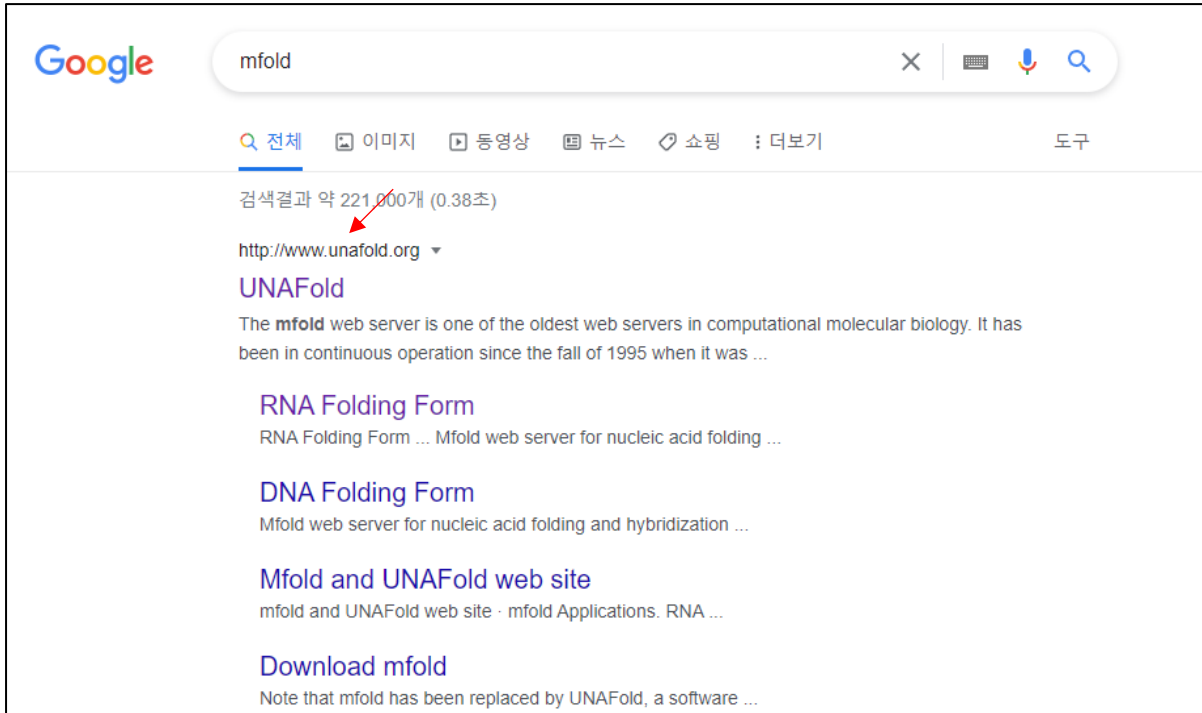
Google search for "rnafold".

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 ...

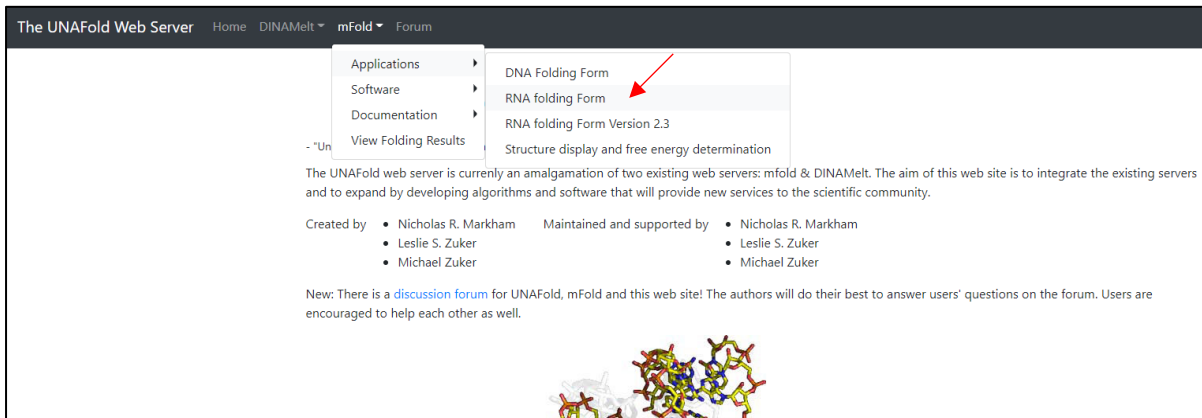
http://rna.tbi.univie.ac.at >
[ViennaRNA Web Services](#)
 Thermodynamic Structure Prediction. **RNAfold** Server ...predicts minimum free energy structures and base pair probabilities from single RNA or DNA sequences.

- Search "mfold" in google and access UNAFold



A screenshot of a Google search for "mfold". The search bar contains "mfold" and the results show approximately 221,000 results. The top result is "UNAFold" with the URL "http://www.unafold.org". Below the title, there is a description: "The mfold web server is one of the oldest web servers in computational molecular biology. It has been in continuous operation since the fall of 1995 when it was ...". There are four sub-links: "RNA Folding Form", "DNA Folding Form", "Mfold and UNAFold web site", and "Download mfold". A red arrow points to the "UNAFold" title.

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



A screenshot of the UNAFold web server interface. The navigation menu is open, showing "Applications" with a sub-menu containing "DNA Folding Form", "RNA folding Form", "RNA folding Form Version 2.3", and "Structure display and free energy determination". A red arrow points to "RNA folding Form". The main content area includes a description of the server, a list of creators (Nicholas R. Markham, Leslie S. Zuker, Michael Zuker), and a list of maintainers (Nicholas R. Markham, Leslie S. Zuker, Michael Zuker). At the bottom, there is a 3D molecular structure visualization.

- 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](#) [Create](#)

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 plot](#)

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](#), [RNAdraw](#), [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](#), [RNAdraw](#), [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](#), [RNAdraw](#), [XRNA ss](#).

- Predict the 3D structure of GFP and compare with X-ray crystallography image
 - Search "colabfold" in google and access ColabFold.

A screenshot of a Google search for "colabfold". The search bar shows "colabfold" with a search icon. Below the search bar, there are filters for "전체" (All), "뉴스" (News), "이미지" (Images), "동영상" (Videos), "쇼핑" (Shopping), and "더보기" (More). The search results show approximately 9,500 results. The first result is "AlphaFold2.ipynb - Colaboratory - Google Colab" with a red arrow pointing to the title. Below it is a link to a Nature article: "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.

A screenshot of the ColabFold AlphaFold2.ipynb notebook interface. The notebook title is "AlphaFold2.ipynb" with a red arrow pointing to it. The interface shows a code cell with the following content:

```

ColabFold: AlphaFold2 using MMseqs2

Easy to use protein structure and complex prediction using AlphaFold2 and Alphafold2-multimer.
Sequence alignments/templates are generated through MMseqs2 and HHsearch. For more details, see
bottom of the notebook, checkout the ColabFold GitHub and read our manuscript. Old versions: v1.0,
v1.1, v1.2, v1.3

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: Making protein folding
accessible to all. Nature Methods. 2022

[2] Input protein sequence(s), then hit Runtime -> Run all
query_sequence: "MSKGEELFTGVVPIVELDGDVNGHKFSVSGEGDATYGKLTCLKFICTTGKLPVPWPTL VTTFSYGVQCFS"
• Use : to specify inter-protein chainbreaks for modeling complexes (supports homo- and hetro-oligomers). For example
PI...SK:PI...SK for a homodimer
jobname: "GFP"
use_amber: 
template_mode: none
• "none" = no template information is used, "pdb70" = detect templates in pdb70, "custom" - upload and search own
templates (PDB or mmCIF format, see notes below)

코드 표시

```

Red arrows in the image point to the notebook title, the "Run all" button, and the "jobname" field.

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

Display 3D structure

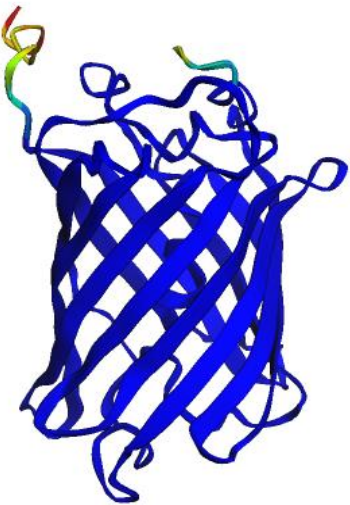
rank_num: 1

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.

