## **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.

Google	ncbi	× 💷 🏓	٩							
(	Q.전체 [글이미지 『 도서 ▶ 동영상 ⊘ 쇼핑 ;더보기		도구							
ł	검색결과 약 608,000,000개 (0.52초)									
t i i i i i i i i i i i i i i i i i i i	nttps://www.ncbi.nlm.nih.gov 🔻									
N N	National Center for Biotechnology Information Welcome to NCBI. The National Center for Biotechnology Information adva realth by providing access to biomedical and genomic information. BLAST	nces science and		미국 국립생물공 < 학정보센터 <sup>회사</sup>	S NCBI					
	BLAST finds regions of similarity between biological sequences.									
	PubMed PubMed® comprises more than 34 million citations for biomedical			미국 국립생물공학정보센터는 미국 보건성 산하 국립 의학도서관의 운영 분야 중 하나이다. NCBI는 1988년 미국 메릴랜드주에 설립되었다. 위키백과						
	Nucleotide The Nucleotide database is a collection of sequences from			<b>창립자: 클</b> 로드 페퍼 <b>창립:</b> 1988년 11월 4일						
	Gene			본사: 미국 메릴랜드 베서스타						
	The .gov means it's official. Federal government websites often end			<b>모회사:</b> 미국 국립 의학 도서관						

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

Search NCBI	human tRNA-	Val	X Se	earch	
Results found in 15 databases					
GENE TRNV – tRNA-Val Homo sapiens (human) Also known as: MTTV Gene ID: 4577 PubMed (7) Genome Data Viewer			Was this helpful?		
Literature		Genes		Proteins	
Bookshelf	33	Gene	214	Conserved Domains	0
MeSH	0	GEO DataSets	0	Identical Protein Groups	0
NLM Catalog	0	GEO Profiles	0	Protein	747
PubMed	191	HomoloGene	0	Protein Family Models	0
PubMed Central	13,754	PopSet	464	Structure	В

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

Search results tems: 1 to 20 of 2	14	<< First < Pre	v Page 1 of 11 Next	> Last >
See also 166 dis	continued or replaced items.			
Name/Gene ID	Description	Location	Aliases	MIM
TRV-CAC1-7 ID: 100189228	tRNA-Val (anticodon CAC) 1-7 [Homo sapiens (human)]	Chromosome 1, NC_000001.11 (149712552149712624, complement)	TRNAV22, TRV- CAC5-1	
D: 4577	mitochondrially encoded tRNA valine [ <i>Homo sapiens</i> (human)]	Chromosome MT, NC_012920.1 (16021670)	MTTV, TRNV	
TRV-AAC1-4 ID: 7239	tRNA-Val (anticodon AAC) 1-4 [ <i>Homo sapiens</i> (human)]	Chromosome 5, NC_000005.10 (181218270181218342, complement)	TRNAV1, TRV, TRV1	18992
D: 100189416	tRNA-Val (anticodon CAC) 3-1 [Homo sapiens (human)]	Chromosome 19, NC_000019.10 (47246354724707, complement)	TRNAV32	61530
TRV-CAC1-6	tRNA-Val (anticodon CAC) 1-6 [Homo sapiens (human)]	Chromosome 6, NC_000006.12 (2653805426538126)	TRNAV21	61530

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 GTITCCGTAGTGTAGTGGTTATCACGTTCGCCTAACACGCGAAAGGTCCCCGGTTCGAAACCGGGCGGAA ACA

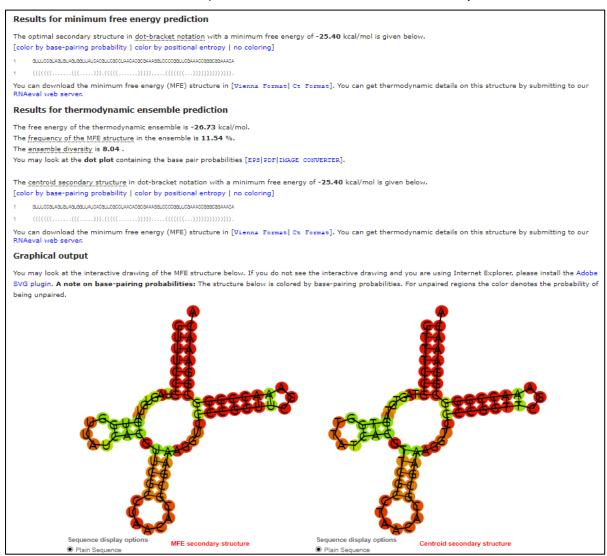
> Search "RNAfold" in google and access RNAfold.

Google	rnafold	X 🔳 🎙 🔍
	Q 전체 ▲ 이미지	도구
	검색결과 약 89,600개 (0.49초)	
	http://rna.tbi.univie.ac.at > cgi-bin RNAWebSuite > R 💌	
	RNAfold web server	
	The <b>RNAfold</b> web server will predict secondary structures of single stranded sequences. Current limits are 7,500 nt for partition function	RNA or DNA
	http://rna.tbi.univie.ac.at 💌	
	ViennaRNA Web Services Thermodynamic Structure Prediction. RNAfold Serverpredicts minimum structures and base pair probabilities from single RNA or DNA sequences.	free energy

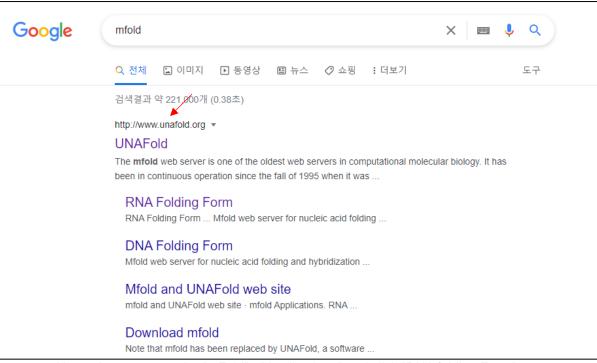
Copy and paste the tRNA sequence to the box, and click "Proceed" for RNA structure prediction.

RNAfold WebServer	1 Enter Input Parameters 2	View Results
	[Home New job	b[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 predicitions.		
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]
(xkC.00005.10: <u>c181218342</u> -181218270 Hose sapiens chronosces 5. <u>(BRC)80.014</u> Primary Assembly GTTICDIAGGENATICACGTICACCTANCLOSEGALAGGTOCOCCEGATOSAA J&		*
P Show constraint folding		
Or upload a file in FASTA format: 파월 선택 사업된 파일 없음		
Fold algorithms and basic options		
minimum free energy (MFE) and partition function go		
minimum free energy (MFE) only		
no GU pairs at the end of helices 🍪		
avoid isolated base pairs 🤫		
P Show advanced options		
Output options		
interactive RNA secondary structure plot (a)		
RNA secondary structure plots with reliability annotation (Partition function folding only)		
Mountain plot 🥹		
Notification via e-mail upon completion of the job (optional): your e-mail		
	Proceed »	

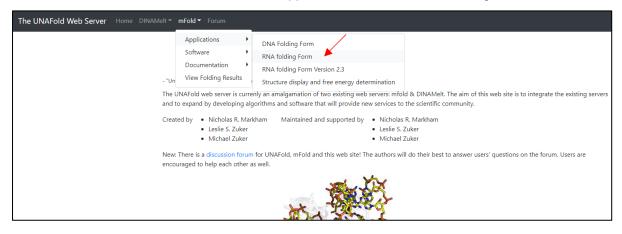
In the result, we can see the optimal secondary structure of tRNA and its free energy. The left one is the structure that has minimum free energy, and the right one is the structure that has minimum total-base pair distance to all structures in the thermodynamic ensemble.



> Search "mfold" in google and access UNAfold

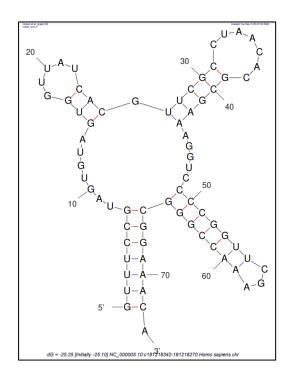


Hover the mouse on the "mFold" - "Applications" and click "RNA folding Form"



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. <b>31 (13)</b> , 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 GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTAACACGCGAAAGGTCCCCGGGTTCGAAACCGGGCGGAA 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
<u>Output</u>
The <i>energy dot plot</i> for NC_000005 10:c181218342-181218270 Homo sapiens chromosome 5, GRCh38 p1. ( <i>Definition</i> ) File formats: <i>Text, PostScript, pdf, png, jpg</i> Computed Structures: ( <i>File Formats</i> ) The computed foldings contain 32 base pairs out of 44 (72.7%) in the <i>energy dot plot</i> .
Extra files: sorted ct file; <i>h-num</i> values; <i>p-num</i> values; <i>log file</i> for main computations.
<u>Download all foldings</u>
Choose <i>zipped</i> file: • or <i>gzipped tar</i> file: • PostScript • Create
View ss-count information
(Definition)(ss-count file)ss value = 1.22 ± 1.23Averaging window1Magnification1Base to magnify about1Plot formatPostScript < View plot
View Individual Structures:
Circular structure Plots ♦ Structure 1 : Initial ΔG = -25.10 kcal mol, (Thermodynamic Details). Different file formats: <i>PostScript, pdf, png, jpg, .ct file, Vienna, RNAML, RnaViz ct, Mac ct, RNAdraw, XRNA ss</i> .
Structure 2 : Initial ΔG = -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 ΔG = -24.50 kcal/mol, (Thermodynamic Details). Different file formats: PostScript, pdf, png, jpg, .ct file, Vienna, RNAML, RnaViz ct, Mac ct, RNAdraw, 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
MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATVGKLTLKFICTTGKLPVPWPTL
VTTFSVGVQCFSRVPDHMKQHDFFKSAMPEGVVQERTIFFKDDGNVKTRAEVKFEGDTLV
NRIELKGIDFKEDGNILGHKLEVNVNSHNVVIMADKQKNGIKVNFKIRHNIEDGSVQLAD
HYQQNTPIGDGPVLLPDNHVLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELVK
```

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

Google	psipred	х 🔳 🎙 ९	
	Q 전체 ▲ 이미지 ▪ 동영상 ▣ 뉴스 ♡ 지도 : 더보기	도구	
	검색결과 약 115,000개 (0.31초)		
	http://bioinf.cs.ucl.ac.uk>psipred ▼ PsiPred - UCL Bioinformatics group 이 페이지에 관한 정보가 없습니다. 이유 알아보기		
	https://bio.tools>psipred *		Park Construction
	PSIPRED - bio.tools		·····································
	The <b>PSIPRED</b> Protein Analysis Workbench unites many available analysis too web based framework. An excellent tool for prediction of secondary	ls into a single	PSIPRED <

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

Submission details	
Protein Sequence	
> <u>spIP42212IGFP_AEQVI</u> Green fluorescent protein OS= <u>Aeguorea victoria</u> OX=6100 <u>GN=GFP</u> PE=1 <u>SV</u> =1 <u>MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTL</u> <u>VTTFSYGVQCFSRVPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV</u>	* *
Help If you wish to test these services follow this link to retrieve a test fasta sequence.	
Job name	
GFP	
Email (optional)	
Email (optional)	
Reset	

In the result, we can separate  $\alpha$ -helix,  $\beta$ -sheet, and coil region.

 $\succ$ 

Sequence Plot																																											
			Г	Sho	ow	psi	pre	d	1			Sh	IOW	me	msa	it			[	Sh	ow	aaty	pes																				
									10								20								30	)							4	10								50	
1	М	S	K	G	EI	ΕI	LF	Т	G	V	V	P 1	L	۷	Е	L	G	D	۷	N	G H	K	F	s١	/ S	G	Е	GE	. 0	D	Α	Т	Y	G	K	L 1	ΓL	. K	F	T.	С	ТТ	50
51	G	Κ	L	P١	VI	P	WF	РТ	L	V	T 1	F	s	Υ	G	v (	Q C	F	s	R	Y P	D	н	M	( Q	н	D	FF	K	s	Α	М	Ρ	E	G١	Y۱	V	Q E	R	т	T.	F F	100
101	К	D	D	G	N	ΥI	K 1	R	А	Е	V	K F	Е	G	D	τI	. V	Ν	R	I. I	EL	К	G	I I	D F	К	Е	D	G N	1.1	L	G	н	κı	LI	E١	YI	N Y	'N	s	н	NV	150
151	Υ	L.	M	A	DI	ĸ	Q	( N	G	1	K۱	/ N	F	К	L.	RH	H N	T.	Е	D	GS	V	Q	L	A D	н	Y	Q	Q N	ΙT	Ρ	$\mathbf{I}_{i}$	G	D	GI	P١	V L	. L	Ρ	D	Ν	ΗY	200
201	L	s	Т	Q	s .	A	LS	к	D	Ρ	NE	ĸ	R	D	н	M١	/ ι	L	Е	F	VТ	A	Α	GI	т	н	G	м	DE	L	Y	к											238
									10								20								30	)							4	10								50	
						otei	n bi	ndin	g			lix tativ -entr				ound	ary						tera	oction	I				insm	neml	l bran tide		elix		G	et i	PNC	3	(	Get	SVG	3	

- 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 <a href="http://www.sbg.bio.ic.ac.uk/phyre2">http://www.sbg.bio.ic.ac.uk/phyre2</a>)

Google	phyre2 X I C								
	Q 전체 🖬 이미지 🗈 동영상 🕮 뉴스 🧷 쇼핑 🕴 더보기 도구								
	검색결과 약 140,000개 (0.19조)								
	도움말: 이 검색을 <b>영어</b> 검색결과로 제한합니다. 언어별 필터링에 관해 자세히 알아보기								
	http://www.sbg.bio.ic.ac.uk>~phyre2 :	Phyre < Phyre <sup>2</sup>							
	PHYRE2 Protein Fold Recognition Server	소프트웨어							
	The PHYRE automatic fold recognition server for predicting the structure and/or function of your protein sequence.								
		영어에서 번역됨 - Phyre 및 Phyre2는 단백질 구조 예							
	https://www.nature.com > nature protocols > protocols	측을 위한 무료 웹 기반 서비스입니다. Phyre는 1500번 이상 인용된 단백질 구조 예측을 위한 가장 인기 있는							
	The Phyrez web portal for protein modeling, prediction and	방법 중 하나입니다. 위키백과(영어)							
	LA Kelley 저술 · 2015 · 9539회 인용 — <b>Phyre2</b> is a web-based tool for predicting and analyzing	원래 설명 보기 🗸							
	protein structure and function. Phyre2 uses advanced remote homology detection	제공 언어: 영어							
	National Institutes of Health (.gov) https://www.ncbi.nlm.nih.gov.y.articles.y.PMC5298202	관련 검색어 4개 이상 항목 더보기							
	The Phyre2 web portal for protein modelling, prediction								
		SWISS- MODELLEF HH-suite IntFOLD							
	National Institutes of Health ( gov) https://pubmed.ncbi.nlm.nih.gov	MODEL 의견 보내기							
	The Phyro2 web portal for protein modeling, prediction and								

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".

Phyre <sup>2</sup>		Subscribe to Phyre at Google Groups Email: Subscribe Visit Phyre at Google Groups J Follow @Phyre2server							
Protein Homology/analogY Recognition Engine V 2.0									
		🔤 🕓 😢 🗹 📦							
Log in (top left-hand corner of this web page) to use our Expert Mode feal Login page $\blacksquare$ .	tures, including batch job submission 🗉 and One-to-One t	hreading <sup>11</sup> . Registering for this is quick and easy on our							
You can use "One-to-One Threading" to model your sequence against you AlphaFold Protein Structure Database	You can use "One-to-One Threading" to model your sequence against your own in-house structures and those from the Protein Data Bank <sup>(1)</sup> , as well as models downloaded from the AlphaFold Protein Structure Database <sup>(1)</sup> .								
If you have more than 5 or 6 sequences to model, it is easier for you (and	better for everyone!) if you use "batch" processing (under	er Expert Mode).							
Please do not use "intensive mode" unless your search using "not most of the time, 'normal mode' will give you the answer you require in a	Please do not use "intensive mode" unless your search using "normal mode" indicates that a single model does not cover most of your sequence. II For most users, most of the time, 'normal mode' will give you the answer you require in a fraction of the time.								
Curren	t Phyre2 server load = 1% (normal running) 🗉								
E-mail address: Optional Job description									
Amino Acid Sequence 🛙	or enter your sequence here								
or upload contents of sequence file	파일 선택 선택된 파일 없음								
or UniProt accession									
	Normal  Intensive Test								
Please tick as appropriate.	NOT for Profit (Commercial)    Other • Phyre Search Reset	1							

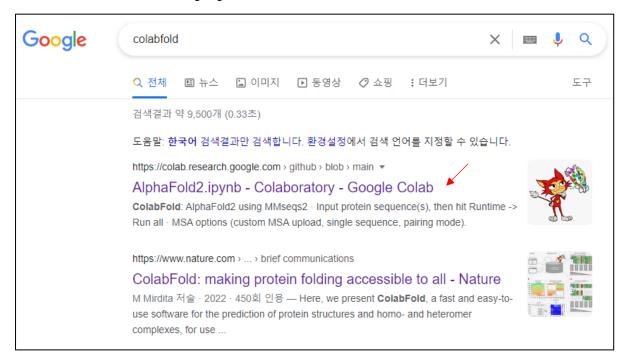
> Wait until to finish their prediction.



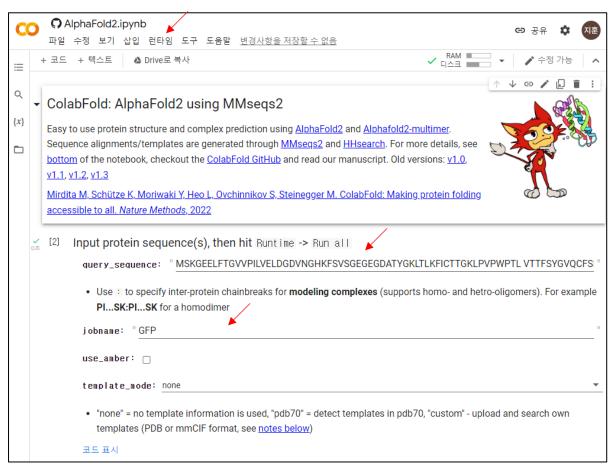
> In the result, we can see the 3D structure of GFP, and some of the data

Phyre <sup>2</sup>	Email hijs502@gmail.com Description GFP Date Thu Nov 16 06:37:30 GMT 2023 Unique 306 DD ccf86e404f8972d4 Sequence MSKGELFTG <u>Download FASTA</u> Job Type <b>normal</b> Job Expiry 30 days
	Top model
	Model (left) based on template <u>c6wvdA_</u> PDB header: membrane protein Chain: A: PDB Molecule: green fluorescent protein, protein jagunal homolog 1 PDBTItle: human jagn1 PDB Entry: PDBe RCSB PDBj - Confidence on decomposition Confidence Confiden
	228 residues ( 96% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template. Division in the single highest scoring template. Interactive 3D view in JSmol For other options to view your downloaded structure offline see the FAQ
Image coloured by rainbow N $\rightarrow$ C terminus	
Image coloured by rainbow N $\rightarrow$ C terminus Model dimensions (Å): X:48.957 Y:55.903 Z:48.782	Confidence: 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

> Search "colabfold" in google and access ColabFold



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



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

