

# **Protein Structure Prediction & Ligand Docking 실습**

김유빈, 정우엽

# 실습에 들어 가기 전에

# Download Chimera

- Go to <https://www.cgl.ucsf.edu/chimera/download.html>

## Current Production Releases

- See the [release notes](#) for a list of new features and other information.
- For [more recent changes](#), use the [snapshot](#) and [daily](#) builds; they are less tested but usually reliable.
- **64-bit Releases:**

| Platform                 | Installer, Size, and Checksum   | Date         | Notes   |
|--------------------------|---|--------------|---|
| Microsoft Windows 64-bit | <a href="#">chimera-1.16-win64.exe</a><br>Size: 152332561 bytes<br>MD5: 9672aa27cc7ea1d6cfe9c8680516c741        | Dec 17, 2021 | <a href="#">Instructions</a><br><a href="#">Documentation</a><br>Runs on Windows 7 or later.      |
| Mac OS X 64-bit          | <a href="#">chimera-1.16-mac64.dmg</a><br>Size: 192170325 bytes<br>MD5: 02cef4e3bf4e2ad5aae44c7104328ade        | Dec 17, 2021 | <a href="#">Instructions</a><br><a href="#">Documentation</a><br>Runs on Mac OS X 10.12 or later. |
| Linux 64-bit             | <a href="#">chimera-1.16-linux_x86_64.bin</a><br>Size: 154080130 bytes<br>MD5: 0167c57d7e24c9b69e04fd0dabc5ce87 | Dec 17, 2021 | <a href="#">Instructions</a><br><a href="#">Documentation</a><br>Compiled on CentOS 5.11.         |

- **32-bit releases are no longer supported.**

# Input & Result File Download

- Go to [https://github.com/seoklab/Public-practice/tree/main/AI-Bio\\_2023](https://github.com/seoklab/Public-practice/tree/main/AI-Bio_2023)

The screenshot shows a GitHub repository named 'Public Practice'. The repository contains a single file, 'README.md'. The main content area includes sections for 'Introduction' and 'Notice'. The 'Introduction' section states: 'This is a repository for public files used in the lectures, presentation etc.' The 'Notice' section states: 'The files in this repository will be permitted to use'. It lists two rules:

1. only for the purpose with permission in the lecture, presentation
2. only for group members or lecture participants with permission in the lecture, presentation

**Download** (1)

If you want to download the specific directory or file, You can use the [DownGit](#) made by [Minhas Kamal](#)

This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).  
Copyright © 2022- Seoul National University Lab of Computational Biology and Biomolecular Engineering.

The screenshot shows a GitHub repository page for 'Public-practice / AI-Bio /'. A red box labeled (2) highlights the browser's address bar, which displays the URL. The repository has a single commit from 'yeopyeob' titled 'AI-Bio input and result upload'. The commit details show two files uploaded: 'AF\_monomer' and 'GalaxyTongDock', both labeled as 'AI-Bio file upload'. The commit was made 1 hour ago.

The screenshot shows the 'DownGit' interface, which is a tool for creating download links for GitHub resources. A red box labeled (3) highlights the input field where a GitHub URL is entered. Below the input field are two buttons: 'Create Download Link' and 'Download'.

# Modeling Target



# Model Protein - 5xra

5XRA

Crystal structure of the human CB1 in complex with agonist AM11542

PDB DOI: <https://doi.org/10.2210/pdb5XRA/pdb>

Classification: SIGNALING PROTEIN

Organism(s): Homo sapiens, Desulfovibrio vulgaris str. Hildenborough

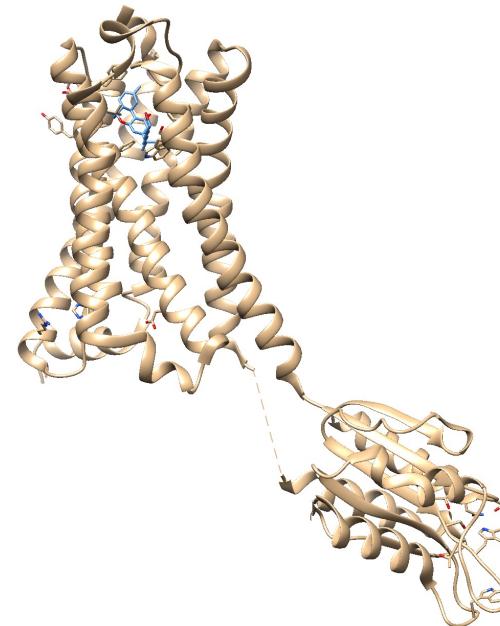
Expression System: Homo sapiens

Mutation(s): Yes ⓘ

Membrane Protein: Yes ⓘ [OPM](#) [PDBTM](#) [MemProtMD](#) [mpstruc](#)

Deposited: 2017-06-08 Released: 2017-07-12

Deposition Author(s): [Hua, T.](#), [Vemuri, K.](#), [Nikas, P.S.](#), [Laprairie, R.B.](#), [Wu, Y.](#), [Qu, L.](#), [Pu, M.](#), [Korde, A.](#), [Shan, J.](#), [Ho, J.H.](#), [Han, G.W.](#), [Ding, K.](#), [Li, X.](#), [Liu, H.](#), [Hanson, M.A.](#), [Zhao, S.](#), [Bohn, L.M.](#), [Makriyannis, A.](#), [Stevens, R.C.](#), [Liu, Z.J.](#)

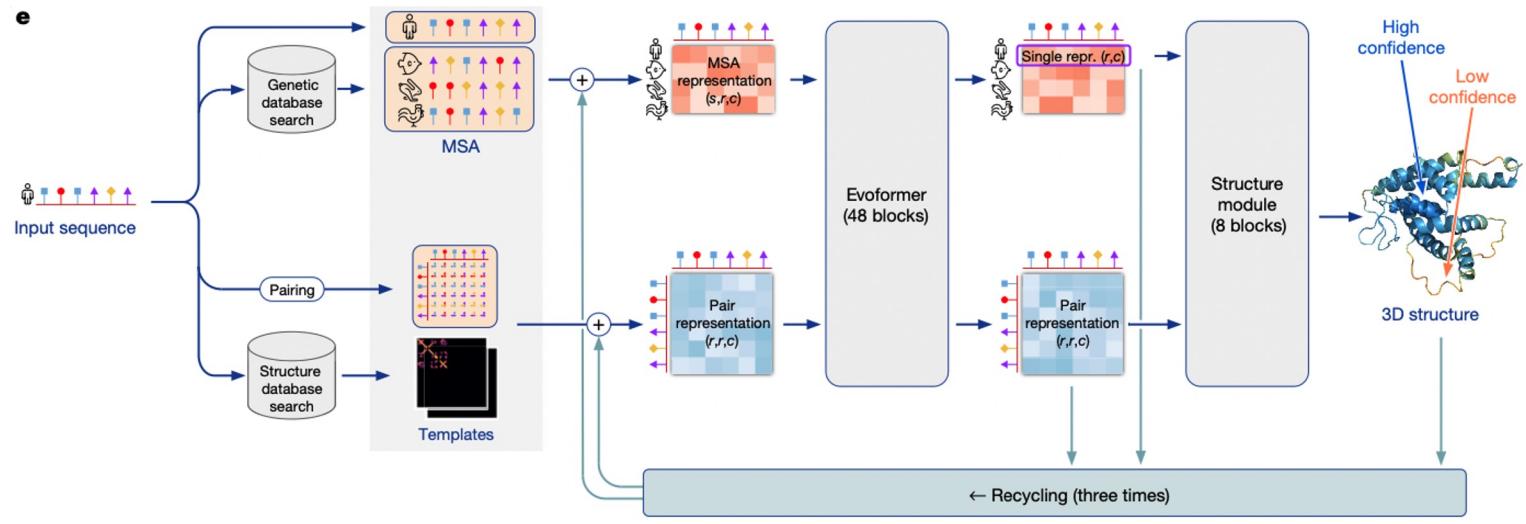


GPCR CB1 + Agonist AM11542

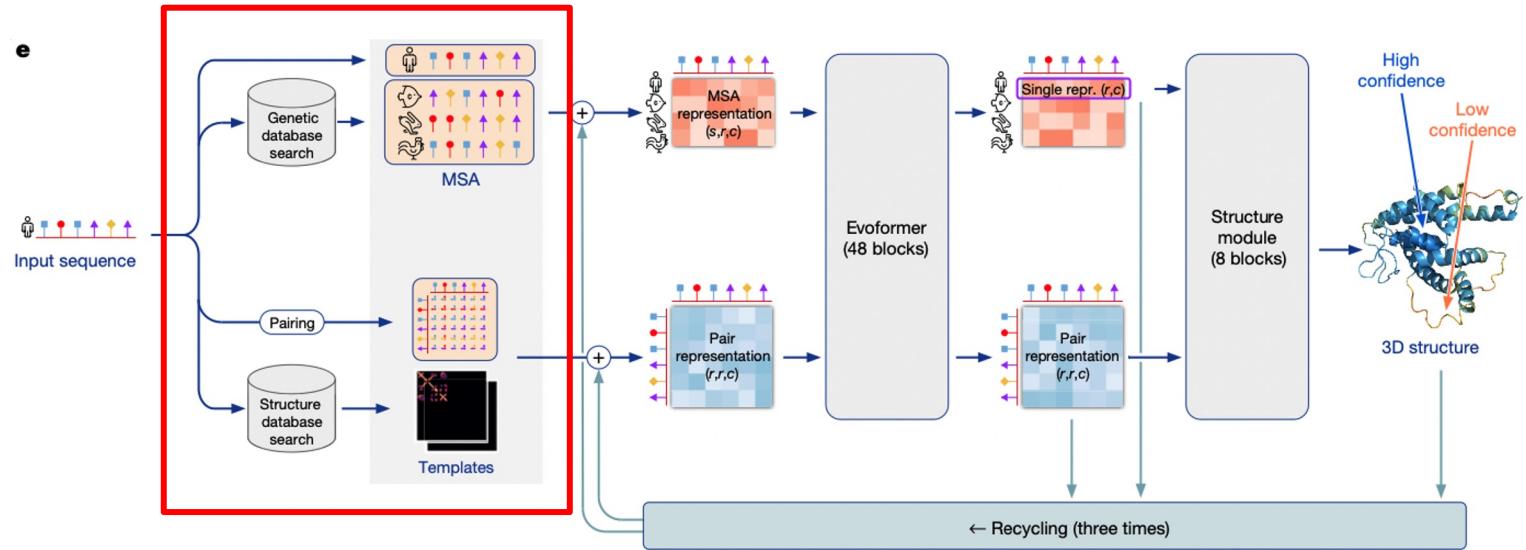
# AF2, AF-multimer

- GPCR and its complex prediction

# What is AlphaFold2, AlphaFold-Multimer?



# What is AlphaFold2, AlphaFold-Multimer?

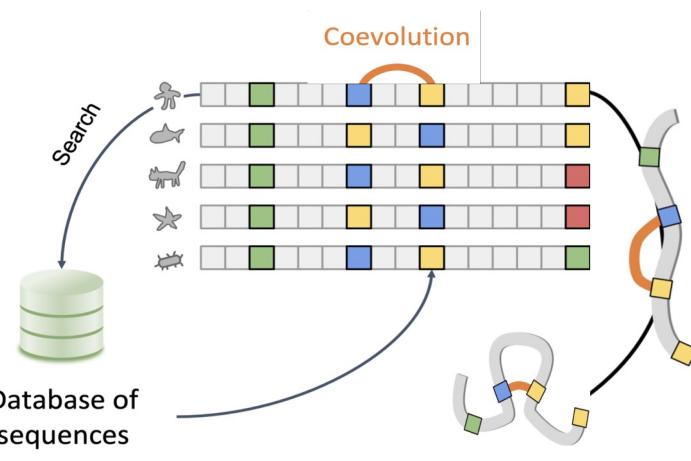


## 1. Build MSA & Get template information (HHblits, HMMer)

# What is AlphaFold2, AlphaFold-Multimer?

## 1. Build MSA & Get template information (HHblits, HMMer)

### Monomer



### Multimer - Paired MSA

#### UniRef30 paired

MQQDDDFQNFVATILESCPDLGGIIGGPRIVLTVALDHDIDESWISLQLIIDYGL  
MSENNNSFENFTVITLESWSWLSOSCISCSKIKLTNEALENVELEEMWSSIIINYSWT  
MNSUSDHEFDVAVILESWSWLSGIGSGSPWPLXLIIDNAVNHGHNLKALLSIVIQYSWT  
MTG---HEDEFVATILESWSWLSGIGSGSKIKLTTEALENVNAAEYELIDRILINYSWT  
DDHQNQFVATILESCPDLGGIIGGPRIVLTVALDHDIDESWISLQLIIDYGL

Paired by species

#### Unpaired

FF---TIPSTPGI---PCUSCGSTIWT  
EHDQSFVAVILESWSWLSGIGSGSPWPL  
GTTQGFVITLLESWSWLSGIGIGACRIP  
IAEFDTATLTELTLNACVSCSPWPK  
MSGISDHFQTVVILSSWLSGIGSGSPWPK  
MSETEAFQDFVITLLESWSWLSGIGSGSPWPK  
MSAVDQFESTILVWISTLIA---PCUSCGSTIWT

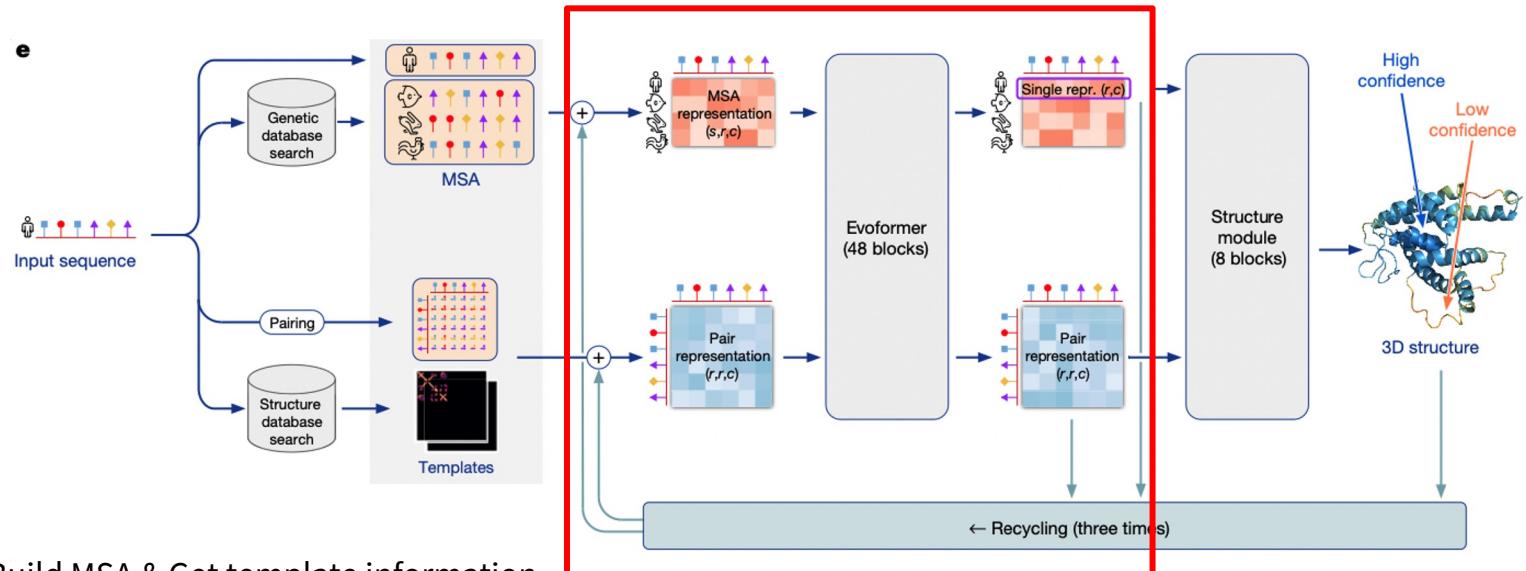
+

LVEISLTDIDYLSWVIAALYSSCKA  
LIEIAISNVSDAEIIQILYSNCKA  
LTDIAVKEENNEDPLISLYSCKA  
LTDIAUWNVSRKSVTUDAYASCKA  
LSTIASLTDIDYLSWVIAALYSSCKA  
MTSICLIVENQYECVGLIQSLTIFRS

*Nat Methods* 19, 679–682 (2022).

*Nature* 596, 583–589 (2021)

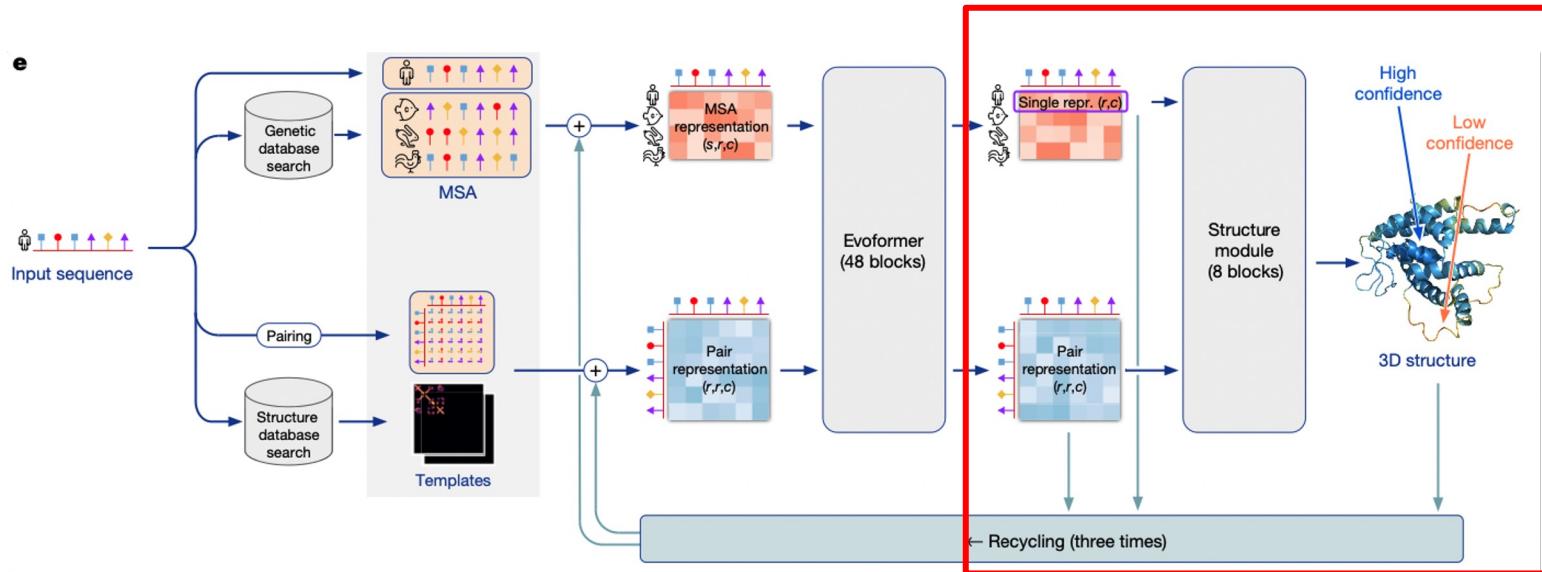
# What is AlphaFold2, AlphaFold-Multimer?



1. Build MSA & Get template information

2. Obtain MSA representation & Pair representation through Evoformer

# What is AlphaFold2, AlphaFold-Multimer?

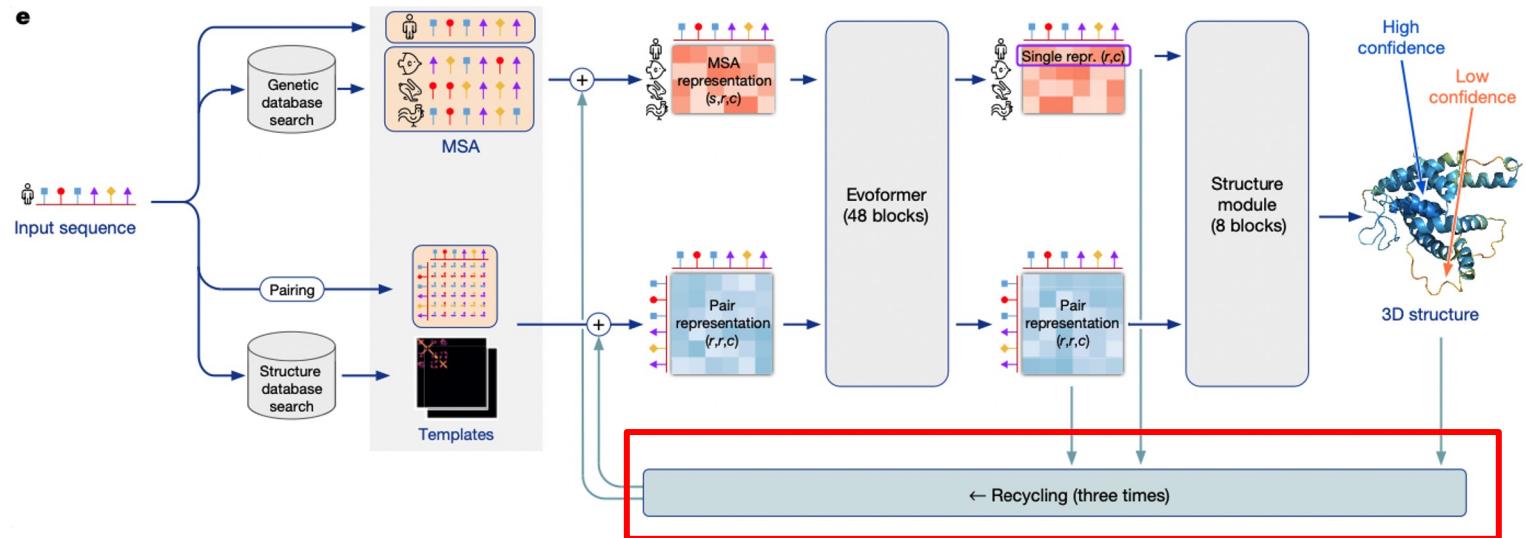


1. Build MSA & Get template information (HHblits, HMMer)

2. Obtain MSA representation & Pair representation through Evoformer

**3. Predict structure using Structure module**

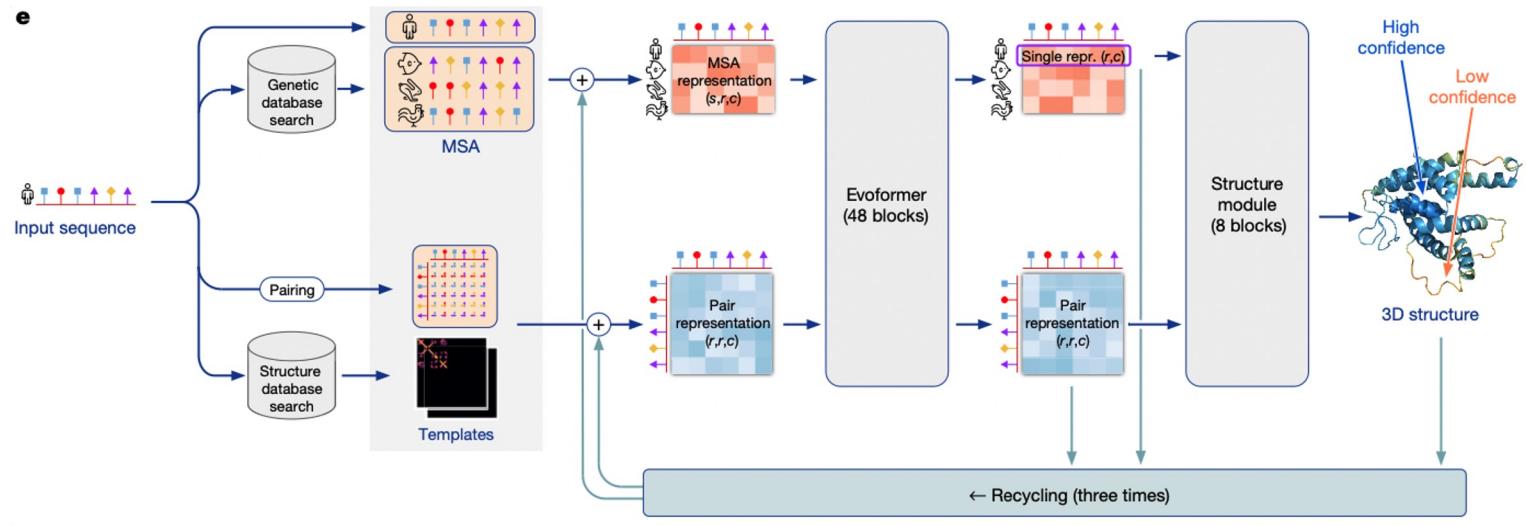
# What is AlphaFold2, AlphaFold-Multimer?



1. Build MSA & Get template information (HHblits, HMMer)
2. Obtain MSA representation & Pair representation through Evoformer
3. Predict structure using Structure module

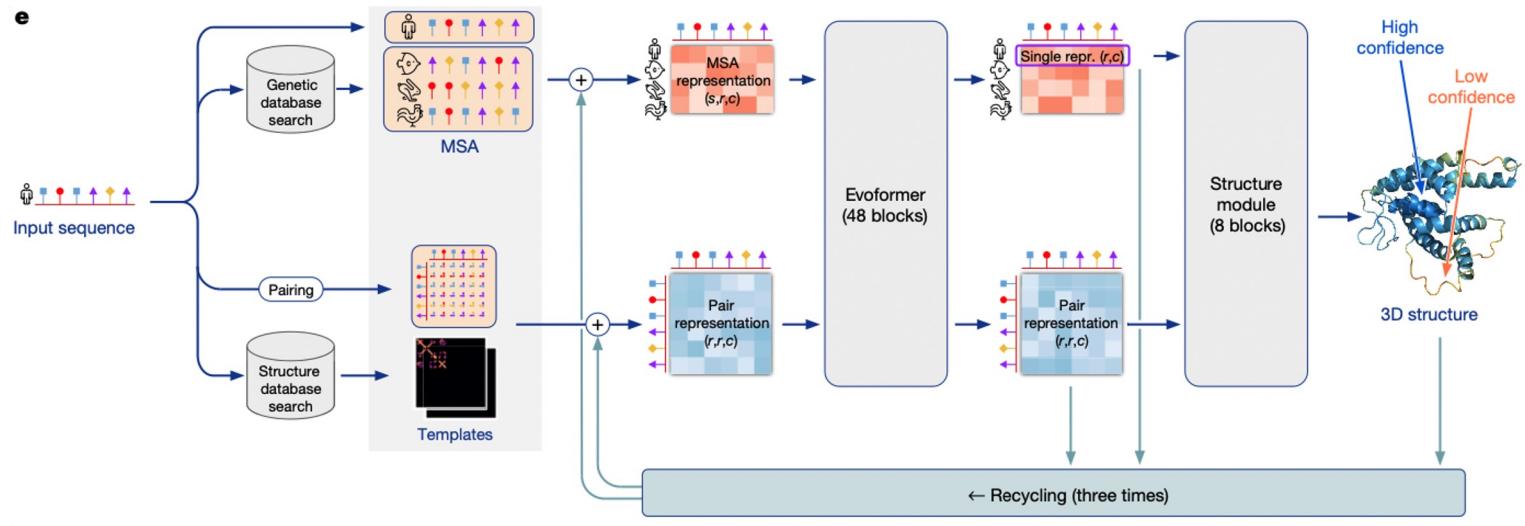
## 4. Iterative Refinement through recycling

# What is AlphaFold2, AlphaFold-Multimer?



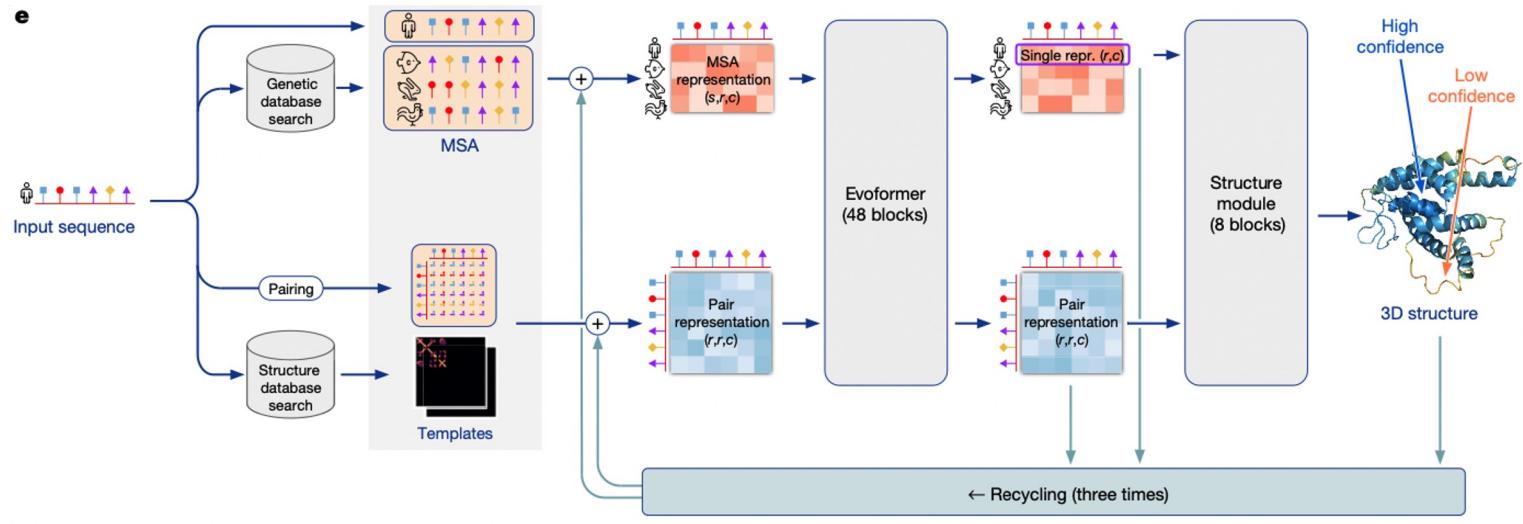
1. Build MSA & Get template information (HHblits, HMMer)
2. Obtain MSA representation & Pair representation through Evoformer
3. Predict structure using Structure module
4. Iterative Refinement through recycling

# What is AlphaFold2, AlphaFold-Multimer?



1. Build MSA & Get template information (HHblits, HMMer) - **several hours, need ~2TB for database**
2. Obtain MSA representation & Pair representation through Evoformer
3. Predict structure using Structure module
4. Iterative Refinement through recycling

# What is AlphaFold2, AlphaFold-Multimer?



1. Build MSA & Get template information (HHblits, HMMer)

2. Obtain MSA representation & Pair representation through Evoformer

3. Predict structure using Structure module

4. Iterative Refinement through recycling

**Need GPUs with a large amount of RAM**

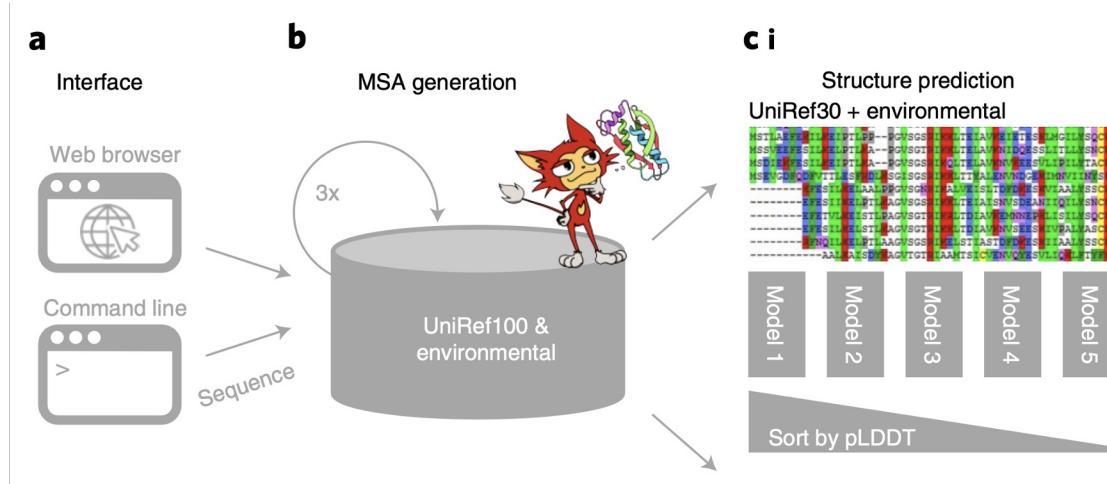
# AF2, AF-multimer 실습

- GPCR and its complex prediction w/ ColabFold

# ColabFold



- <https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>
- Build MSA by MMseqs2 (40-60 folds Faster)
- Speed up structure predictions by avoiding recompilation and adding an early stop criterion (~90 folds faster)



# Model Protein - 5xra

## 5XRA

Crystal structure of the human CB1 in complex with agonist AM11542

PDB DOI: <https://doi.org/10.2210/pdb5XRA/pdb>

Classification: SIGNALING PROTEIN

Organism(s): *Homo sapiens, Desulfovibrio vulgaris str. Hildenborough*

Expression System: *Homo sapiens*

Mutation(s): Yes ⓘ

Membrane Protein: Yes ⓘ OPM PDBTM MemProtMD mpstruc

Deposited: 2017-06-08 Released: 2017-07-12

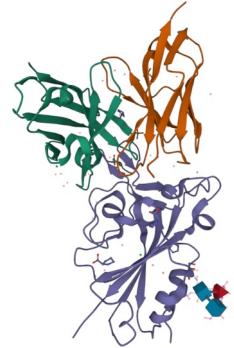
Deposition Author(s): Hua, T., Vemuri, K., Nikas, P.S., Laprairie, R.B., Wu, Y., Qu, L., Pu, M., Korde, A., Shan, J., Ho, J.H., Han, G.W., Ding, K., Li, X., Liu, H., Hanson, M.A., Zhao, S., Bohn, L.M., Makriyannis, A., Stevens, R.C., Liu, Z.J.



Modeling GPCR in CB1 5XRA with ColabFold!

GPCR CB1 + Agonist AM11542

# Input Protein sequence

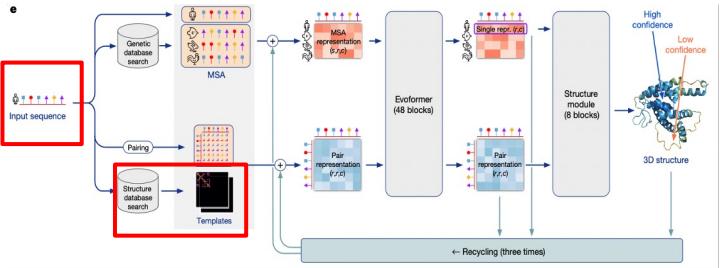


- **Input Files : /Public-practice/AI-Bio\_2023/AF\_monomer/CB1.fa**
  - **Fasta File of CB1 (Protein sequence file)**
  - **RCSB Database for proteins (<https://www.rcsb.org/>) (Protein crystal structure file)**

[https://github.com/seoklab/Public-practice/tree/main/AI-Bio\\_2023/AF\\_monomer](https://github.com/seoklab/Public-practice/tree/main/AI-Bio_2023/AF_monomer)

```
# CB1.fa
> chain_A
ENFMDIECFMVLNPSQQLAIAVSLTLLGTFTVLENLLVLCVILHSRSLRCRPSYHFIGSLAVADLLGSVIFVYSFIDFH
VFHRKDSRNVFLFKLGGVTASFTASVGSLFLAAIDRYISIHRLPLAYKRIVTRPKAVVAFCLMWTTIAIVIAVLPLLGWNC
EKLQSVCSDFPHIDKTYLMFWIGVVSVLLLFIYAYMYILWKAHSHAVAKALIVYGSTTGNTETYTAETIARELADAGY
EVDSRDAASVEAGGLFEGFDLVLLGCSTWGDDSIELQDDFIPLFDSLEETGAQGRKVACFGCGDSSWEYFCGAV
DAIEEKLNKGAEIVQDGLRIDGDPRARRDIVWAHDVRAIPDQARMDIELAKTLVLILVLIICWGPLLAIMVYDV
FGKMNKLIKTVFAFCMSLCLLNSTVNPIIYALRSKDLRHAFRSMFPS
```

# Input Protein sequence



Input protein sequence(s), then hit Runtime -> Run all

query\_sequence: "ENFMIDECFMVLNPSQLAIAVSLTGTFTVLENLLVLCVILHSRSLRCRPSYHFIGSLAVADLLGSVIFVYSFIDFHVFHRKDSRNVLFLKLGVTASFTAVGSLFLAAIDRYISI"

- Use : to specify inter-protein chainbreaks for **modeling complexes** (supports homo- and hetro-oligomers). For example **PI...SK:PI...SK** for a homodimer

jobname: "CB1"

num\_relax: 5

- specify how many of the top ranked structures to relax using amber

template\_mode: pdb100

- none = no template information is used. pdb100 = detect templates in pdb100 (see [notes](#)). custom - upload and search own templates (PDB or mmCIF format, see [notes](#))

코드 표시

Chain이 여러 개인 경우  
각 chain 사이를 : 로 구분

jobname (ex. CB1)

Constrained relaxation of  
structures using Amber force  
field  
(stereochemical violation 제거)

Template\_mode:

- i) **None** - no template information is used
- ii) **pdb100** - detect templates in pdb100 (representative structure database based on the pairwise structure similarity by FoldSeek)
- iii) **custom** - upload & search own templates

# MSA options

▶ MSA options (custom MSA upload, single sequence, pairing mode)

**msa\_mode:** mmseqs2\_uniref\_env

**pair\_mode:** unpaired\_paired

- "unpaired\_paired" = pair sequences from same species + unpaired MSA, "unpaired" = separate MSA for each chain, "paired" - only use paired sequences.

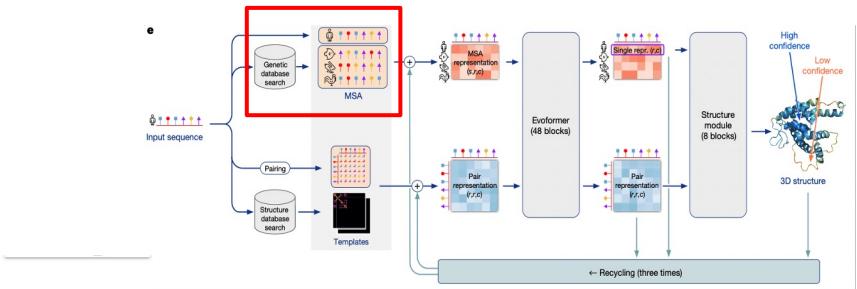
코드 표시

## MSA mode:

- i) MMseqs2 (UniRef + Environmental) - use UniRef sequences + Environmental sequences
  - ii) MMseqs2 (UniRef) - only use UniRef sequences
  - iii) single\_sequence - only use single sequence
  - iv) custom - upload own MSA

## Pair mode:

- i) **unpaired + paired** - pair sequences from same species + unpaired MSA
  - ii) **unpaired** - separate MSA for each chain
  - iii) **paired** - only use paired sequences



## UniRef30 paired

MQQDDDFQNTVATELGGPFDGCGIGGGPPIRGGTIVLALNDIDICGICLICLNNI  
MSNNSNSFENITDLETSVPSQSCISCSGSPRIMVLTALNLENVLEKQVLSNSST  
MHSUDSHEVDYVAILSSFLWLGSCISCSGSPRIMVLTALNLENVLEKQVLSNSST  
MIG---HEDFTAVLSSFLWLGSCISCSGSPRIMVLTALNLENVAAEYVLSNSST  
---DHDNFVATLSSFLWLGSCISCSGSPRIMVLTALNLENVLEKQVLSNSST

## Unpaired

# Advanced Settings

## Advanced settings

`model_type: auto`

### Model type: AlphaFold2 or AlphaFold-Multimer

- if `auto` selected, will use `alphafold2_ptm` for monomer prediction and `alphafold2_multimer_v3` for complex prediction. Any of the mode\_types can be used (regardless if input is monomer or complex).

`num_recycles: 3`

- if `auto` selected, will use `num_recycles=20` if `model_type=alphaFold2_multimer_v3`, else `num_recycles=3`.

`recycle_early_stop_tolerance: auto`

- if `auto` selected, will use `tol=0.5` if `model_type=alphaFold2_multimer_v3` else `tol=0.0`.

`pairing_strategy: greedy`

### Pairing strategy : MSA pairing strategy

- `greedy` = pair any taxonomically matching subsets, `complete` = all sequences have to match in one line.

## Sample settings

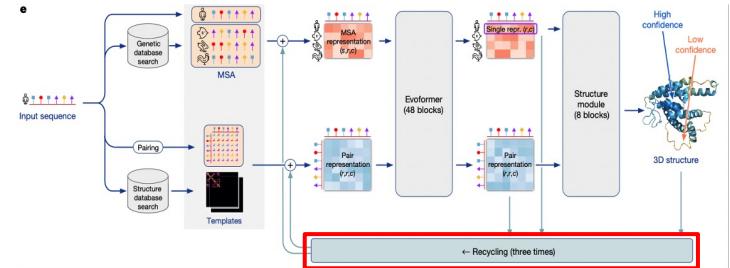
- enable dropouts and increase number of seeds to sample predictions from uncertainty of the model.
- decrease `max_msa` to increase uncertainty

`max_msa: auto`

### Sampling settings : When want to get more diverse structures

`num_seeds: 1`

`use_dropout:`



# Run prediction!

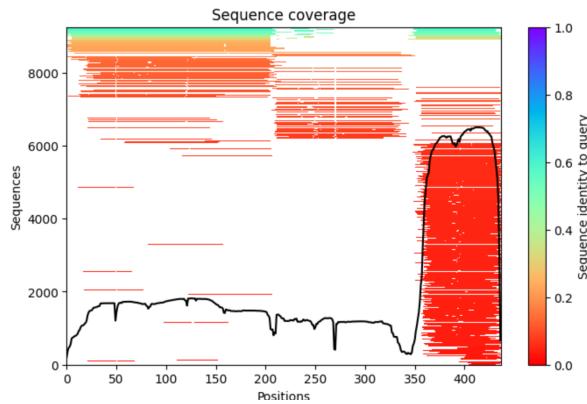
## Run Prediction

```
display_images: ✓  
코드 표시  
...  
2023-10-20 08:01:14.962 Downloading alphafold1d weights to : -i 100% [██████████] 3.47G/3.47G [00:24:00:00, 150MB/s]  
2023-10-20 08:01:14.964 Unable to initialize backend 'room': NOT_FOUND: Could not find registered platform with name: "room". Available platform names are: '  
2023-10-20 08:01:20, 883 Running on GPU  
2023-10-20 08:01:21, 093 Found 8 citations for tools or databases  
2023-10-20 08:01:21, 091 Query 1/1: CBL9e1c (length 437)  
PENDING: 0% | 0/150 [elapsed: 00:00 remaining: 01:02:11] 2023-10-20 08:01:21,822 Sleeping for 10s. Reason: PENDING  
RUNNING: 1% | 1/150 [elapsed: 00:11 remaining: 02:40] 2023-10-20 08:01:38,533 Sleeping for 7s. Reason: RUNNING  
RUNNING: 11% | 17/150 [elapsed: 00:19 remaining: 02:29] 2023-10-20 08:01:40,249 Sleeping for 6s. Reason: RUNNING  
RUNNING: 15% | 23/150 [elapsed: 00:25 remaining: 02:22] 2023-10-20 08:01:46,953 Sleeping for 8s. Reason: RUNNING  
RUNNING: 21% | 31/150 [elapsed: 00:34 remaining: 02:11] 2023-10-20 08:01:55,694 Sleeping for 10s. Reason: RUNNING  
RUNNING: 27% | 41/150 [elapsed: 00:44 remaining: 01:59] 2023-10-20 08:02:02,393 Sleeping for 10s. Reason: RUNNING  
RUNNING: 34% | 51/150 [elapsed: 00:54 remaining: 01:52] 2023-10-20 08:02:10,778 Sleeping for 10s. Reason: RUNNING  
RUNNING: 41% | 61/150 [elapsed: 01:06 remaining: 01:35] 2023-10-20 08:02:27,758 Sleeping for 6s. Reason: RUNNING  
RUNNING: 45% | 67/150 [elapsed: 01:13 remaining: 01:30] 2023-10-20 08:02:34,497 Sleeping for 8s. Reason: RUNNING  
RUNNING: 50% | 75/150 [elapsed: 01:22 remaining: 01:21] 2023-10-20 08:02:43,157 Sleeping for 7s. Reason: RUNNING  
RUNNING: 55% | 82/150 [elapsed: 01:29 remaining: 01:14] 2023-10-20 08:02:50,878 Sleeping for 8s. Reason: RUNNING  
2023-10-20 08:03:19,357 Sequence 0 found templates. ['3cqz_A', '3x10_A', '7v3z_A', '5xra_A', '7ddz_A', '7k15_A', '6i11_A', '7fco_A', '5u09_A', '6kqj_A', '7w  
2023-10-20 08:03:19,357 Sequence 0 found templates. ['3cqz_A', '3x10_A', '7v3z_A', '5xra_A', '7ddz_A', '7k15_A', '6i11_A', '7fco_A', '5u09_A', '6kqj_A', '7w
```

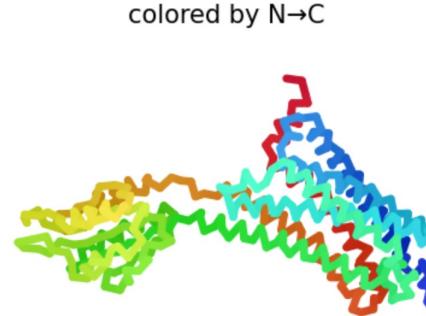
## Log : AlphaFold running log

- Template information
- MSA setting information
- Scoring information : pLDDT & pTM value

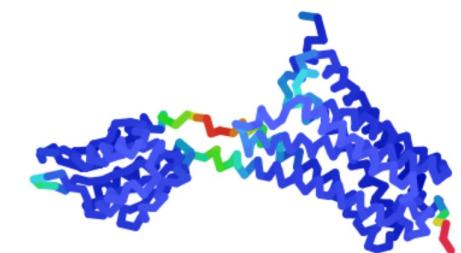
etc



Sequence coverage : MSA information visualization



Colored structure : Visualization with reliability etc



# What is pLDDT, pTM ?

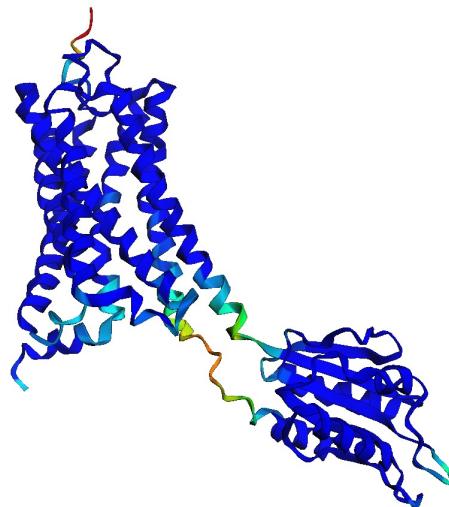
- **pLDDT** (predicted local distance difference test )
  - i. predict the per-residue lDDT-Ca score
  - ii. local error metric
  - iii. value from 0~100 (uncertain ~ certain)
- **pTM** (predicted TM score)
  - i. whether the model is confident in overall domain packing
  - ii. Value from (0-1] (not matched ~perfectly matched)

*Bioinformatics* 29(21): 2722-2728 (2013)

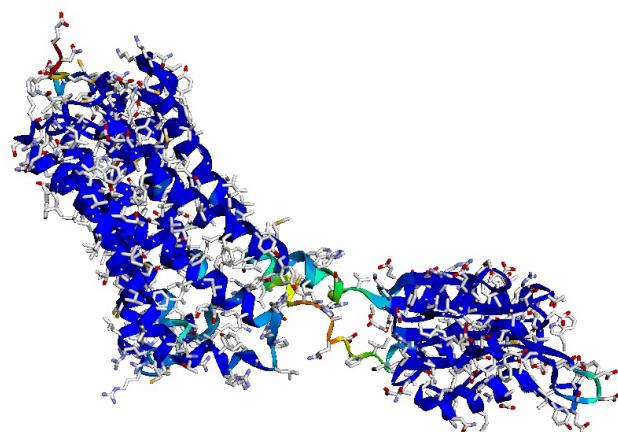
PROTEINS: Structure, Function, and Bioinformatics 57:702–710 (2004)

# Visualization of predicted structures

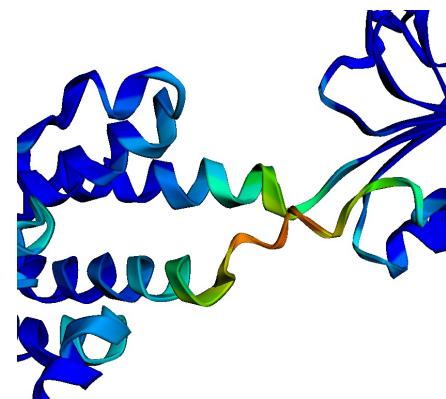
Show side chain X



Show side chain O

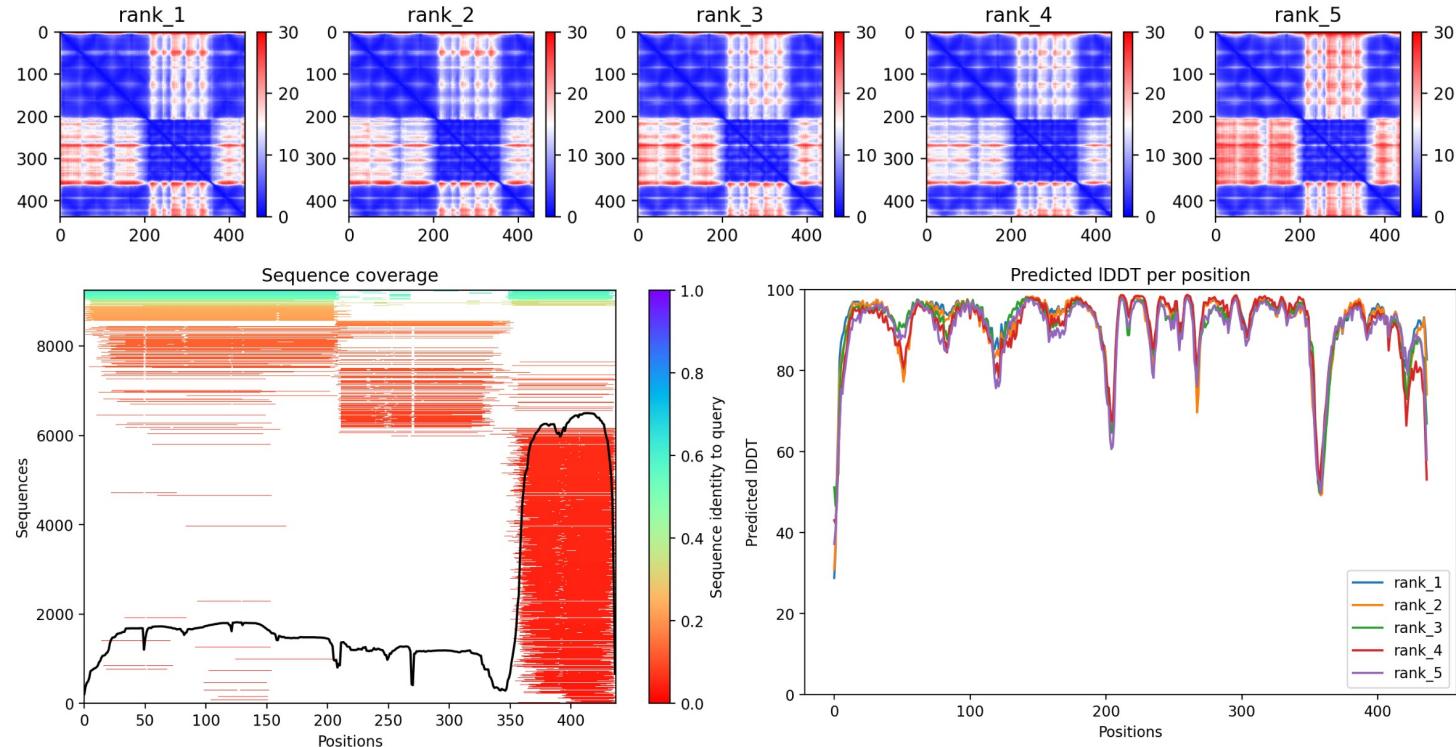


Focusing on the loop



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

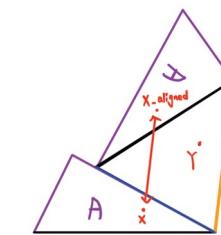
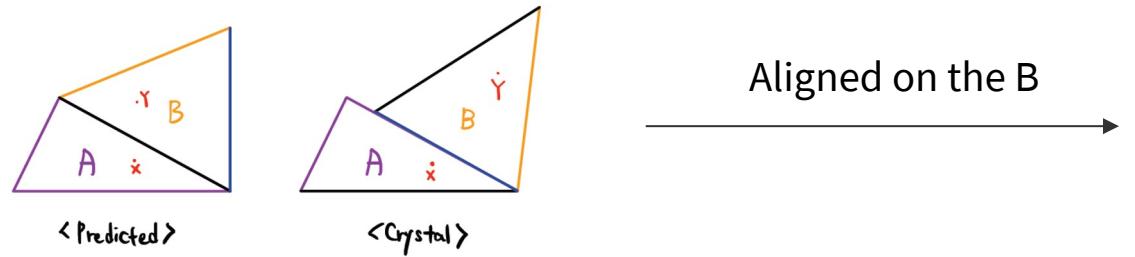
# Plots for predicted structures



# What is PAE ?

- **PAE** (predicted aligned error)

- For every pair  $(x, y)$  of residues in the structure, PAE is calculated by AlphaFold's estimate of position error at residue  $x$ , while the predicted and true structures are aligned on residue  $y$ .
- If the relative position of two domains is confidently predicted, the PAE values will be low (less than 5 $\text{\AA}$ ) for pairs of residues from each domain.



# Multimer case - 6yx9

6YX9

Cryogenic human adiponectin receptor 2 (ADIPOR2) at 2.4 Å resolution determined by Serial Crystallography (SSX) using CrystalDirect

PDB DOI: 10.2210/pdb6YX9/pdb

Classification: MEMBRANE PROTEIN

Organism(s): Homo sapiens

Expression System: Drosophila melanogaster

Mutation(s): No

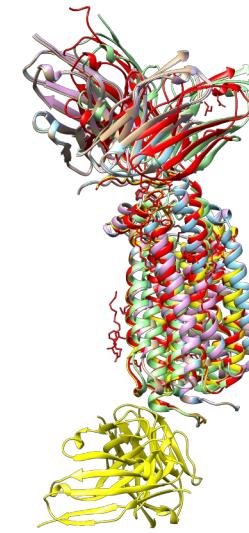
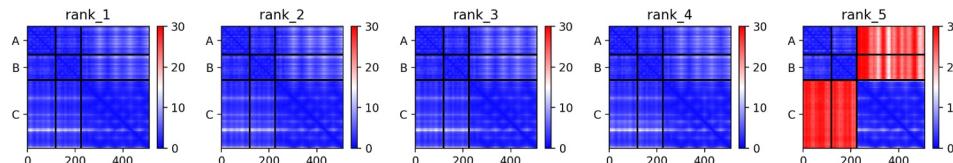
Membrane Protein: Yes

PDBTM

MemProtMD

mpstruc

PAE score for 6yx9 - A: Heavy chain, B: Light chain, C: antigen



Crystal structure - red

Ranked #5 structure - yellow

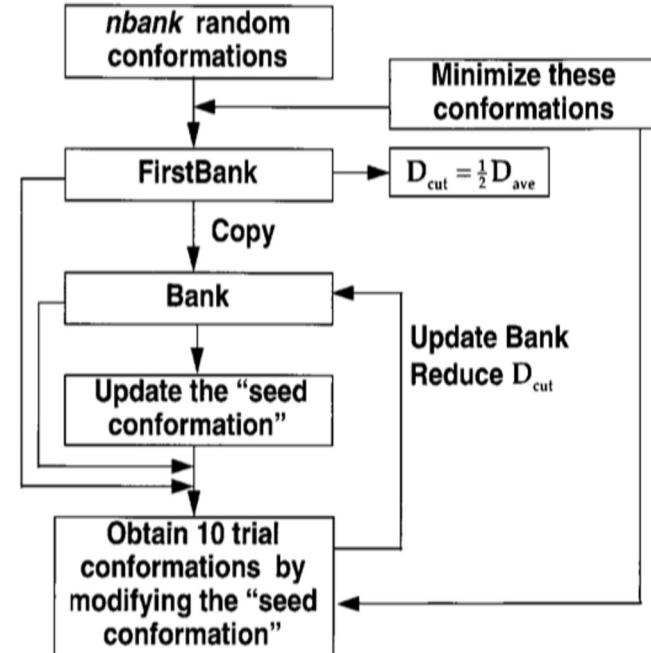
- PAE score can be a criterion of whether the binding geometry is well predicted !

# GalaxyDock 실습



# GalaxyDock : What it is

- Protein-Ligand Docking tool based on CSA
  - CSA : Genetic algorithm + Simulated Annealing
- Based on GalaxyDock BP2 Score
  - Hybrid of physics-based, empirical, knowledge-based
- 6+K Degrees of freedom of Ligand
  - Fixed Ring conformation, Rigid Receptor Protein
  - Degrees of freedom
    - Translation + Rotation + Torsion angle



# GalaxyDock : Preparing input

- Receptor : Crystal structure or Predicted structure by AF2 etc. → .pdb format
- **Ligand**

- Prepare SMILES of the target molecule

```
# lig.smi  
c1(O)cc(cc2OC(C)(C)[C@H]3[C@H](c12)CC(=CC3)C)C(CCCCCCBr)(C)C
```

- Convert SMILES to .mol2 using Corina (<https://mn-am.com/products/corina/>) : 1D→3D

```
corina -i t=smiles lig.smi -o t=mol2 >> lig.mol2
```

- Preprocess mol2 file : attach hydrogen and assign partial charge with Chimera command

```
del H          # delete Hydrogen      addh          # attach Hydrogen  
addcharge all method gas           # assign partial charge
```

# GalaxyDock : Run & Analysis

- Let's Run!! with GalaxyWeb

<https://galaxy.seoklab.org/>

## Input file

/Public-practice/AI-Bio\_2023/GalaxyDock/ lig.mol2

/Public-practice/AI-Bio\_2023/GalaxyDock/CB1\_AF1.pdb

## binding site

78A,94A,180A,382A,402A

**GalaxyDockWEB**

Given a protein receptor structure and a set of ligand structures, protein-ligand complex structures are predicted by the GalaxyDock protein-ligand docking program.

**User Information**

Job name

E-mail address (Optional)

**Input protein and ligand structures**

PDB File  
(≤1000 AA)  
 파일 선택 선택한 파일 없음  
Protein Structure File (allowed file extensions: pdb, txt)

Ligand File  
(≤150 atoms per ligand,  
up to 500 ligands)  
 파일 선택 선택한 파일 없음  
Ligand Structure Files (allowed file extensions: mol2, pdb, xyz)  
Note: Ligand structure with stereochemically wrong topology might results in inaccurate docking. (e.g. 2D-projected structure of non-planar ligands)

**Binding pocket residues**

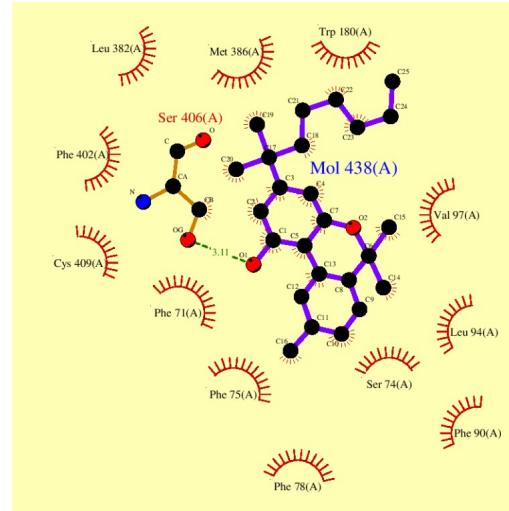
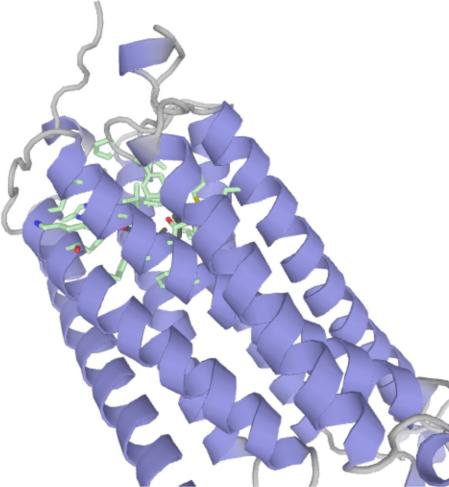
Binding pocket residues  
(≤10 res)  
 Residue numbers should follow the input PDB file residue numbering with chain ID.  
Up to 10 residue numbers can be submitted in integers separated by commas  
(example: 51A,64C,78B).

**Submit**

# GalaxyDock : Run & Analysis

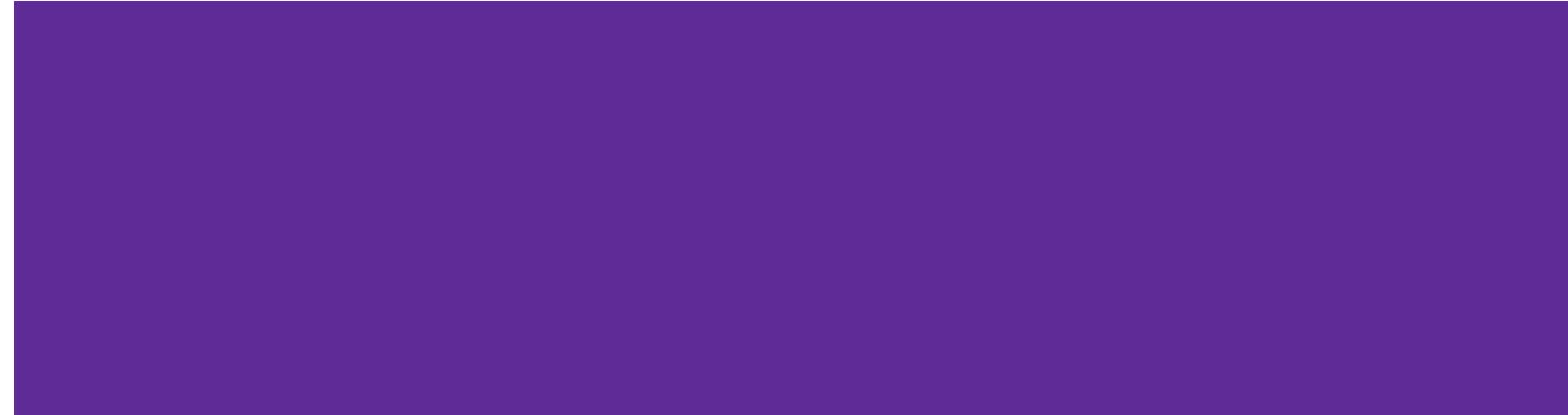
- Result Page

[https://galaxy.seoklab.org/cgi-bin/report\\_DOCK.cgi?key=28ad5065eadfb75fc7b3f21b4848910b](https://galaxy.seoklab.org/cgi-bin/report_DOCK.cgi?key=28ad5065eadfb75fc7b3f21b4848910b)



|     |   |     |     |
|-----|---|-----|-----|
| RES | A | 71  | PHE |
| RES | A | 74  | SER |
| RES | A | 75  | PHE |
| RES | A | 78  | PHE |
| RES | A | 90  | PHE |
| RES | A | 94  | LEU |
| RES | A | 97  | VAL |
| RES | A | 98  | THR |
| RES | A | 101 | PHE |
| RES | A | 168 | ILE |
| RES | A | 169 | PHE |
| RES | A | 172 | ILE |
| RES | A | 177 | LEU |
| RES | A | 180 | TRP |
| RES | A | 382 | LEU |
| RES | A | 386 | MET |
| RES | A | 402 | PHE |
| RES | A | 406 | SER |
| RES | A | 409 | CYS |

# Chimera 실습



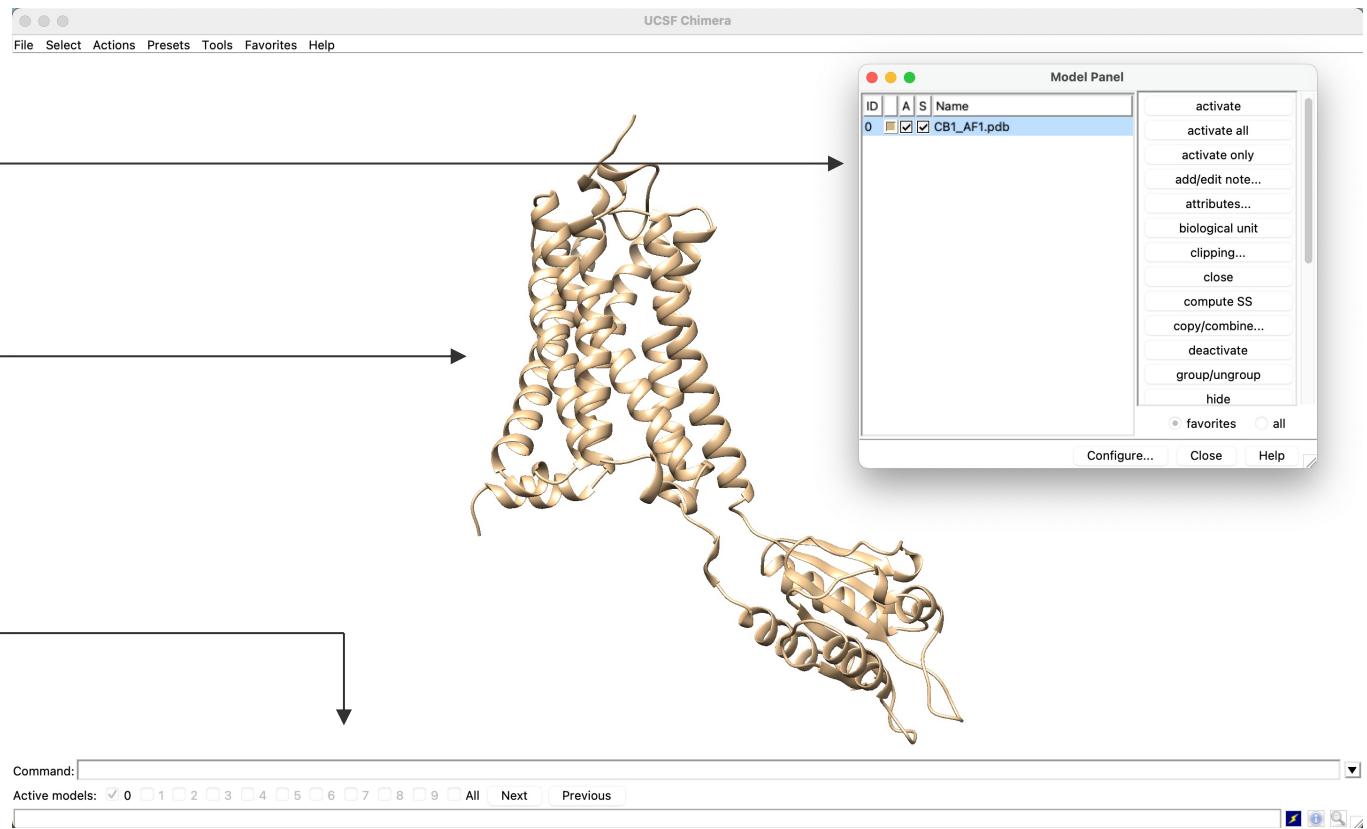
# Visualization of predicted structures with Chimera

When open .pdb

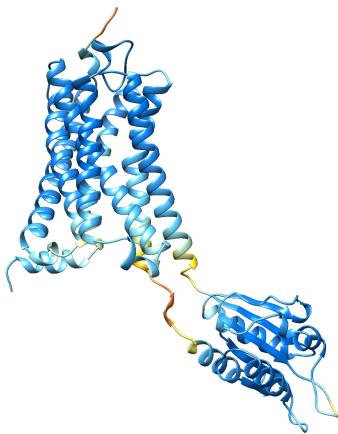
Opened files  
show, active etc

3D structure  
회전, 병진 등 가능

Command line  
3D structure 조작



# Visualization of predicted structures with Chimera

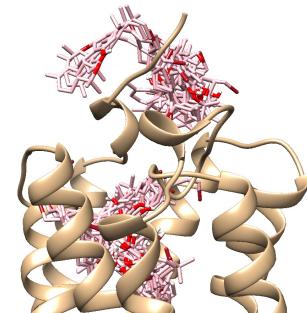


**Let's open the CB1\_AF1.pdb**

- /Public-practice/AI-Bio\_2023/GalaxyDock/CB1\_AF1.pdb

**When want to see pLDDT of AlphaFold model,**

“ range bfactor, 50 #F08253 70 #FADA4D 90 #7EC9EF 100 #1B57CE “



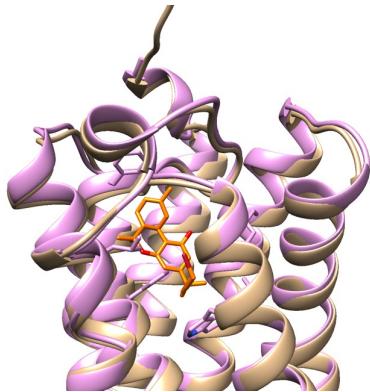
**Let's open the GD\_result.mol2**

- /Public-practice/AI-Bio\_2023/GalaxyDock/GD\_result.mol2

**Too much color, CB1 into basic color and ligand into pink, hetero atom into each color**

“ col tan #1 “ → “ col pink #2 “ → “ col byhet “ ( col [color] [target] # = model, :. = chain, : = residue )

# Visualization of predicted structures with Chimera



Let's open original complex, complex.pdb

- /Public-practice/AI-Bio\_2023/GalaxyDock/complex.pdb

Let's align the complex to the model structure, color it

" mm #0 #2 alg sw " ( mm [ref] [tar] alg sw )

→ " sel #2:8D3 " → " col orange sel " → " col byhet " → " ~sel "

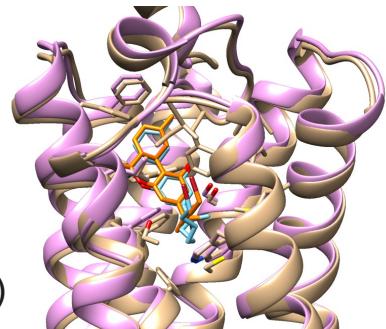
( sel [tar] / #N:R = Residue R in model N, :R.C = Residue in chain C / ~ : not )

To compare the model and crystal, delete all docked ligand except top1  
and show side chain of contacting residues

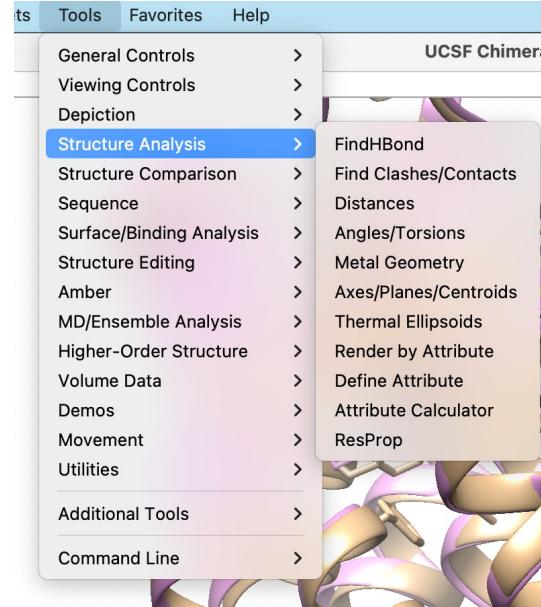
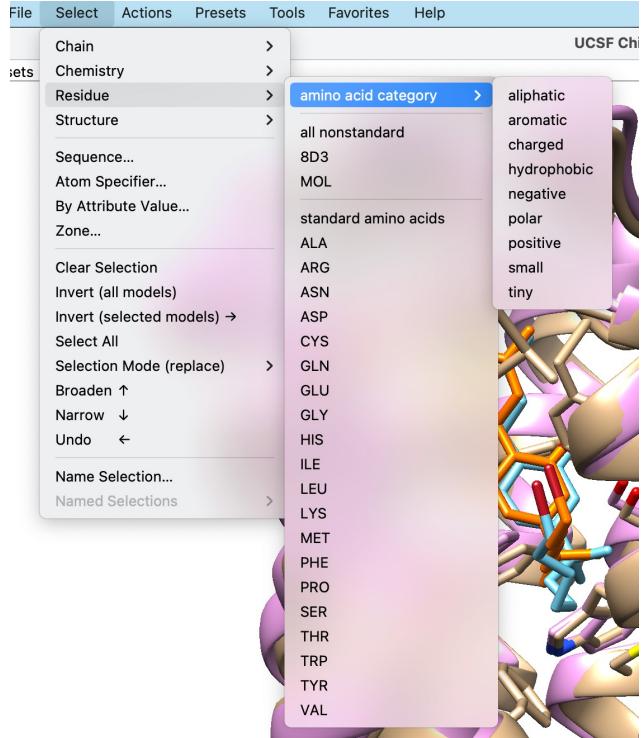
" sel #1:2-50 " → " del sel "

→ "sel :71:74:75:78:90:94:97:98:101:168:169:172:177:180:382:386:402:406:409 "

→ " ~sel #1 #2 " → " di sel " → " del H " ( di [tar] / 연속 = 쉼표 등으로 구분 없이 )

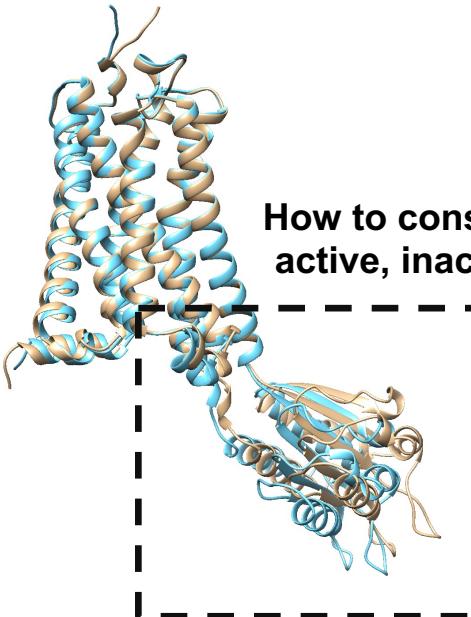


# Visualization of predicted structures with Chimera



**Many other functions**  
<https://www.cgl.ucsf.edu/chimera/docindex.html>

# Limitation of AlphaFold



Input protein sequence(s), then hit Runtime -> Run all

query\_sequence: "PIAQIHLILEGRSDEQKETLIREVSEAIRSLDAPLTSVRVITEMAKGHFGIGGELASK"

• Use : to specify inter-protein chainbreaks for **modeling complexes** (supports homo- and hetero-oligomers). For example PI...SK:PI...SK for a homodimer

jobname: test

num\_relax: 0

• specify how many of the top ranked structures to relax using amber

template\_mode: custom

• none = no template information is used. pdb100 = detect templates in pdb100 (see [notes](#)). custom - upload and search own templates (PDB or mmCIF format, see [notes](#))

코드 표시

파일 선택, 선택한 파일 없음 Cancel upload

## Using custom templates

To predict the structure with a custom template (PDB or mmCIF formatted): (1) change the `template_mode` to "custom" in the execute cell and (2) wait for an upload box to appear at the end of the "Input Protein" box. Select and upload your templates (multiple choices are possible).

- Templates must follow the four letter PDB naming with lower case letters.
- Templates in mmCIF format must contain `_entity_poly_seq`. An error is thrown if this field is not present. The field `_pdbx_audit_revision_history.revision_date` is automatically generated if it is not present.
- Templates in PDB format are automatically converted to the mmCIF format. `_entity_poly_seq` and `_pdbx_audit_revision_history.revision_date` are automatically generated.

If you encounter problems, please report them to this [issue](#).

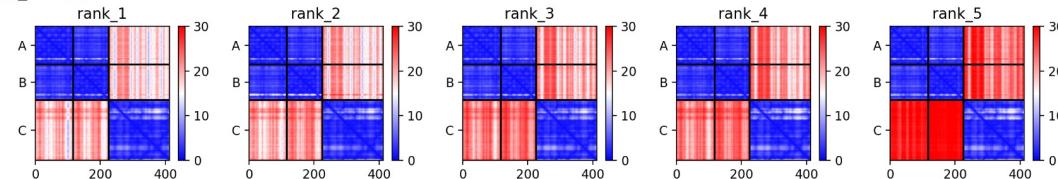
If we use only inactive form templates, then it could predict inactive form!

# Limitation of AlphaFold

**Inaccurate prediction of complex,  
especially antibody-antigen complex**

coevolution information is limited  
due to the unique generation mechanism of antibodies

Plots for 7bem\_89674



→ **Ab-initio docking method, other deep learning based methods are being developed**

( And also, many other deep learning based ligand docking method are being developed )

# **Thank you for listening**