

# Protein Structure Prediction & Visualization

1<sup>st</sup> December, 2023  
Seongeun Kim

# Contents

Protein  
3D  
Structure

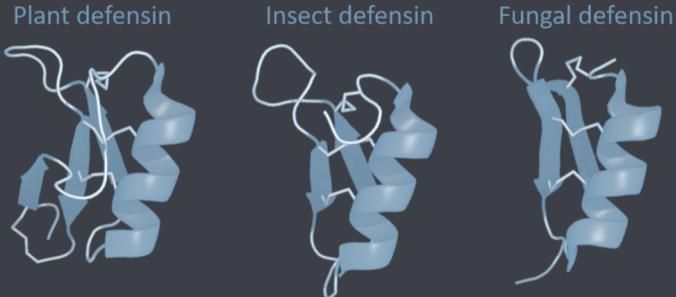
AlphaFold  
Prediction

Visualization  
Tutorial

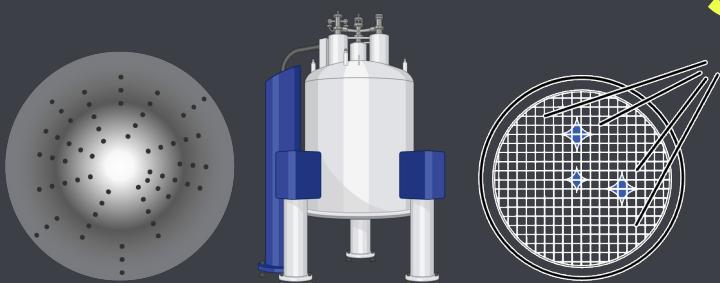
Disordered  
Protein  
Interaction

All-in-one  
Notebook

# Function



<https://en.wikipedia.org/wiki/Defensin>



# Determination

# Archived

RCSSB PDB  
PROTEIN DATA BANK

# Protein Tertiary Structure



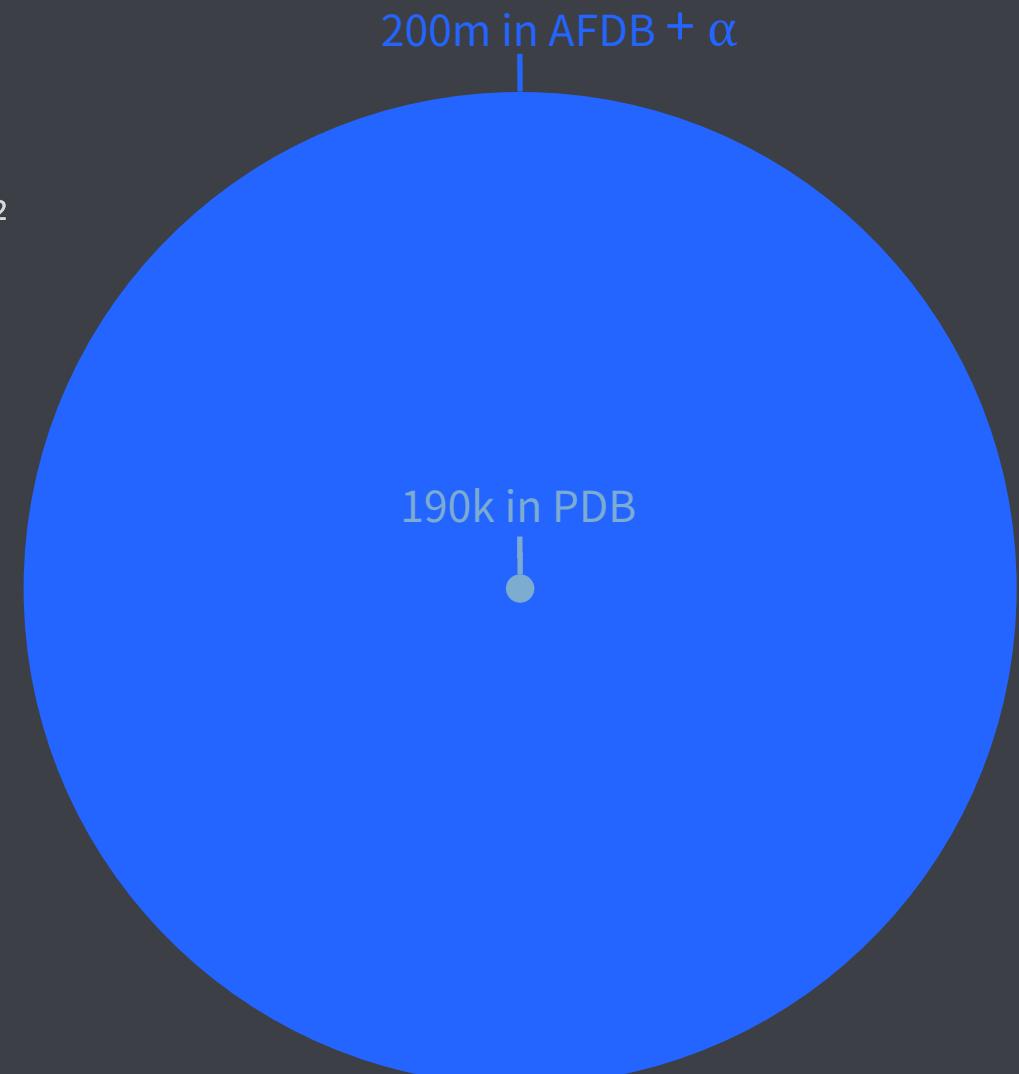
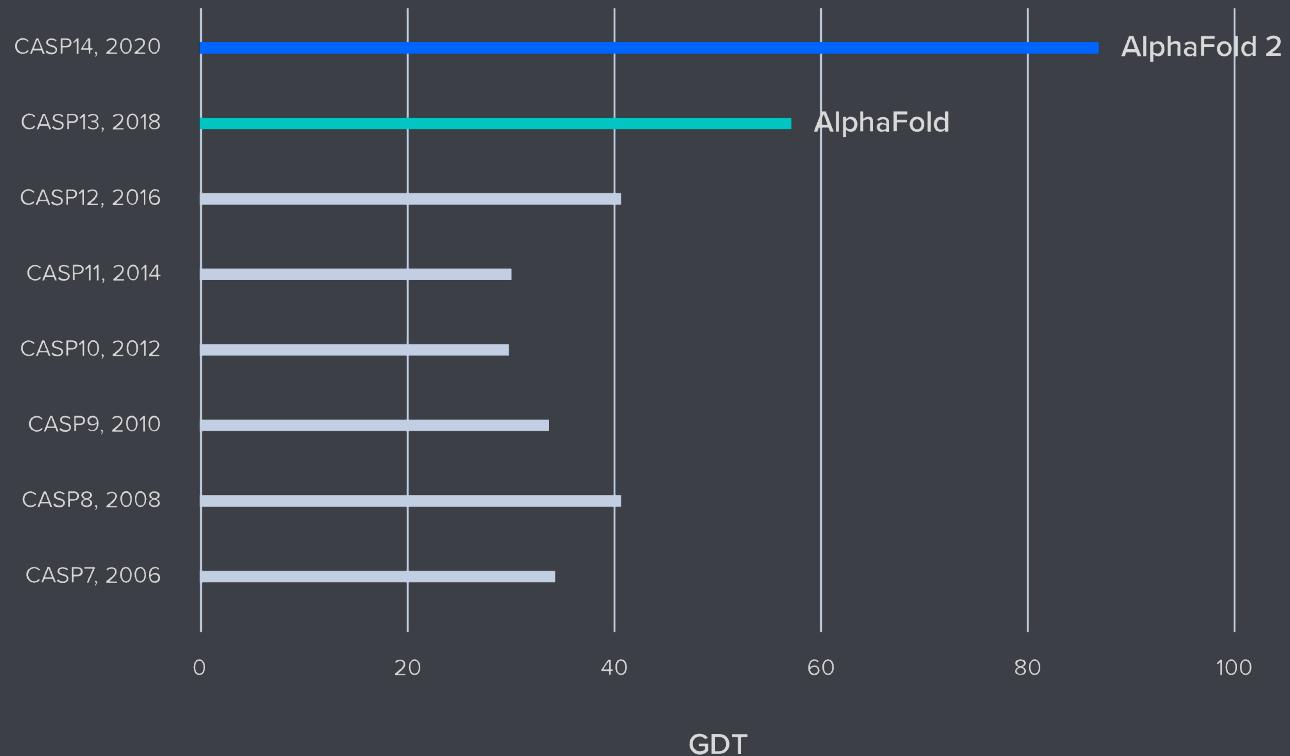
coordinates

ATOM	1	N	HIS	A	1	49.668	24.248	10.436	1.00	25.00	N
ATOM	2	CA	HIS	A	1	50.197	25.578	10.784	1.00	16.00	C
ATOM	3	C	HIS	A	1	49.169	26.701	10.917	1.00	16.00	C
ATOM	4	O	HIS	A	1	48.241	26.524	11.749	1.00	16.00	O
ATOM	5	CB	HIS	A	1	51.312	26.048	9.843	1.00	16.00	C
ATOM	6	CG	HIS	A	1	50.958	26.068	8.340	1.00	16.00	C
ATOM	7	ND1	HIS	A	1	49.636	26.144	7.860	1.00	16.00	N
ATOM	8	CD2	HIS	A	1	51.797	26.043	7.286	1.00	16.00	C
ATOM	9	CE1	HIS	A	1	49.691	26.152	6.454	1.00	17.00	C
ATOM	10	NE2	HIS	A	1	51.046	26.090	6.098	1.00	17.00	N
ATOM	11	N	SER	A	2	49.788	27.850	10.784	1.00	16.00	N
ATOM	12	CA	SER	A	2	49.138	29.147	10.620	1.00	15.00	C
ATOM	13	C	SER	A	2	47.713	29.006	10.110	1.00	15.00	C
ATOM	14	O	SER	A	2	46.740	29.251	10.864	1.00	15.00	O
ATOM	15	CB	SER	A	2	49.875	29.930	9.569	1.00	16.00	C
ATOM	16	OG	SER	A	2	49.145	31.057	9.176	1.00	19.00	O
ATOM	17	N	GLN	A	3	47.620	28.367	8.973	1.00	15.00	N
ATOM	18	CA	GLN	A	3	46.287	28.193	8.308	1.00	14.00	C
ATOM	19	C	GLN	A	3	45.406	27.172	8.963	1.00	14.00	C

# Format

# Structure can also solved by Modelling

Median Free-Modelling Accuracy



<https://www.nature.com/articles/d41586-020-03348-4>

# Protein Structure Visualization

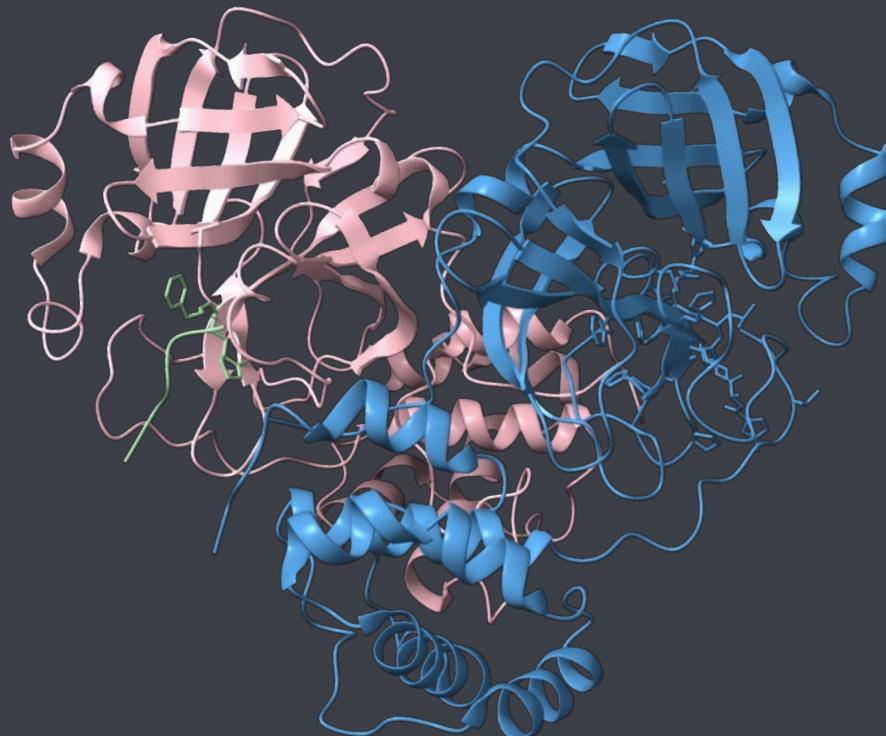
PyMOL  
**UCSF ChimeraX**  
Jmol  
Rasmol  
...

- UCSF ChimeraX  
<https://www.rbvi.ucsf.edu/chimerax/>
- Coda Page  
[https://coda.io/d/ChimeraX-Visualization\\_d5ja4-0cGJ0/How-to-use-Command-Script\\_suhxy#\\_luhKm](https://coda.io/d/ChimeraX-Visualization_d5ja4-0cGJ0/How-to-use-Command-Script_suhxy#_luhKm)
- GitHub (Today's)  
[https://github.com/rachelse/lab\\_231201](https://github.com/rachelse/lab_231201)

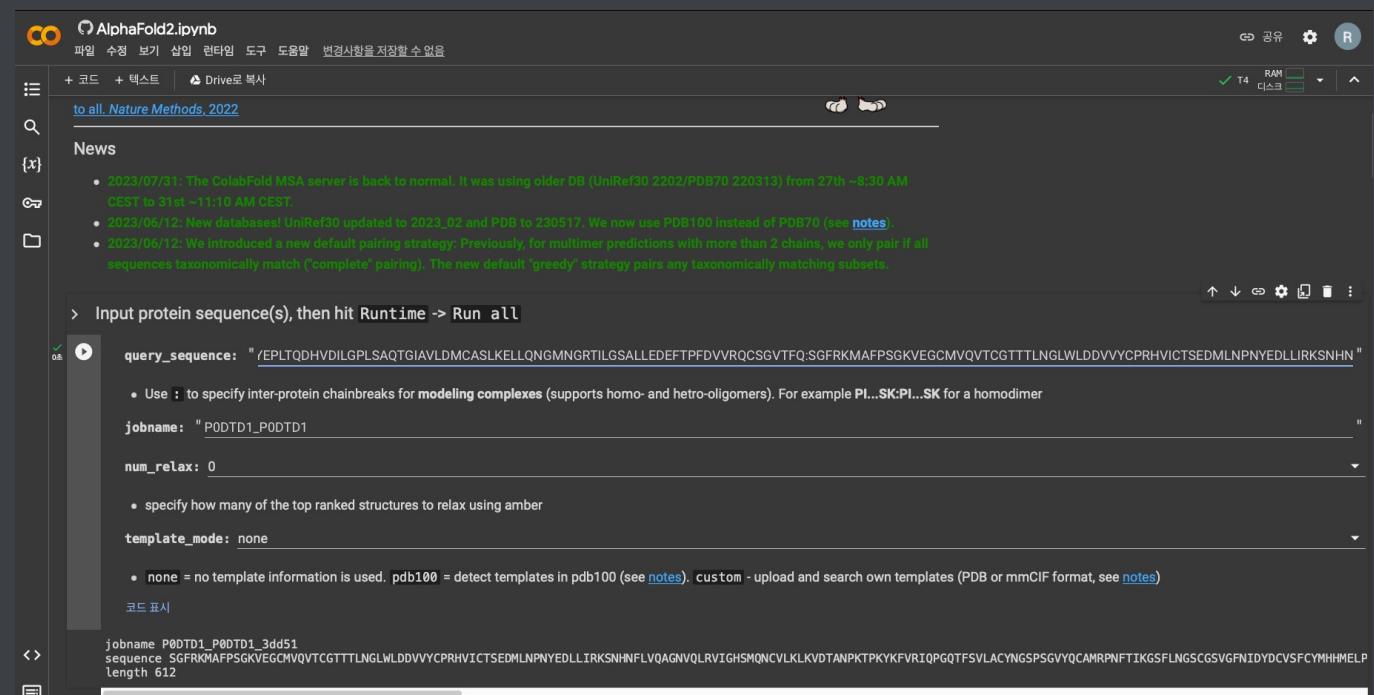
# Today's Protein, Coronavirus Proteases

6lu7

Cut the polyproteins into functional pieces  
Homodimer form two active sites



## Multimer prediction with ColabFold



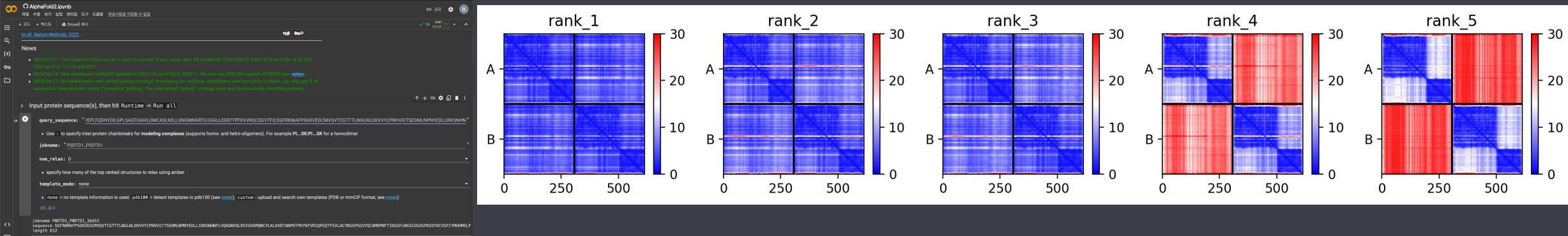
The screenshot shows the AlphaFold2.ipynb notebook interface in Google Colab. The code cell contains the following input parameters:

```
query_sequence: "/EPLTQDHVDILGPLSAQTGIAVLDMCASLKEELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQ:SGFRKMAFPSKGVEGCMVQVTCGTTLNLWLDDVVYCPRHICTSEDMLNPNYEDLLIRKSNNH"
jobname: "P0DTD1_P0DTD1"
num_relax: 0
template_mode: none
```

Notes from the news section:

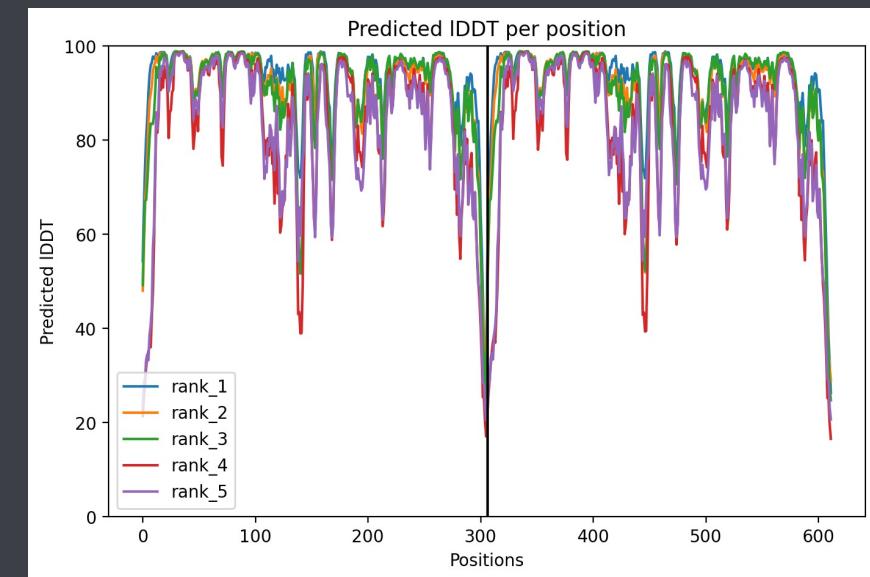
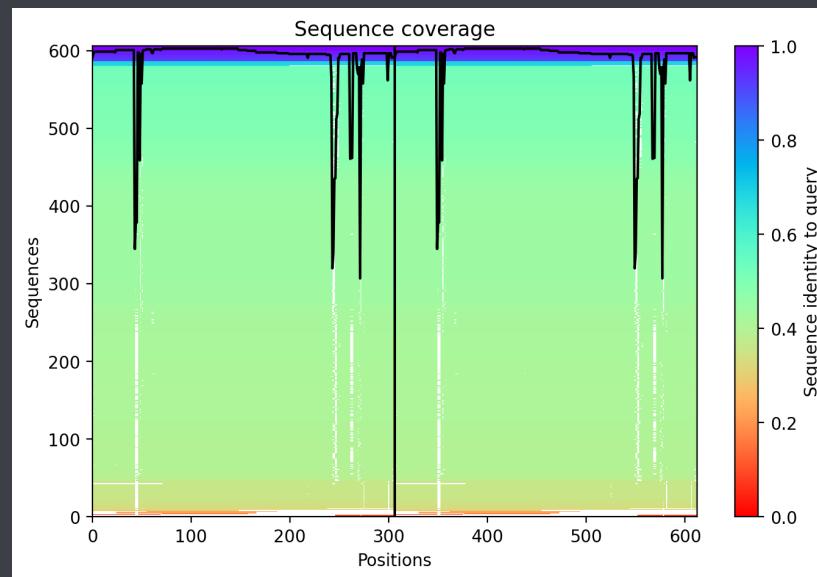
- 2023/07/31: The ColabFold MSA server is back to normal. It was using older DB (UniRef30 2202/PDB70 220313) from 27th ~8:30 AM CEST to 31st ~11:10 AM CEST.
- 2023/06/12: New databases! UniRef30 updated to 2023\_02 and PDB to 230517. We now use PDB100 instead of PDB70 (see notes).
- 2023/06/12: We introduced a new default pairing strategy: Previously, for multimer predictions with more than 2 chains, we only pair if all sequences taxonomically match ('complete' pairing). The new default 'greedy' strategy pairs any taxonomically matching subsets.

# ColabFold prediction output

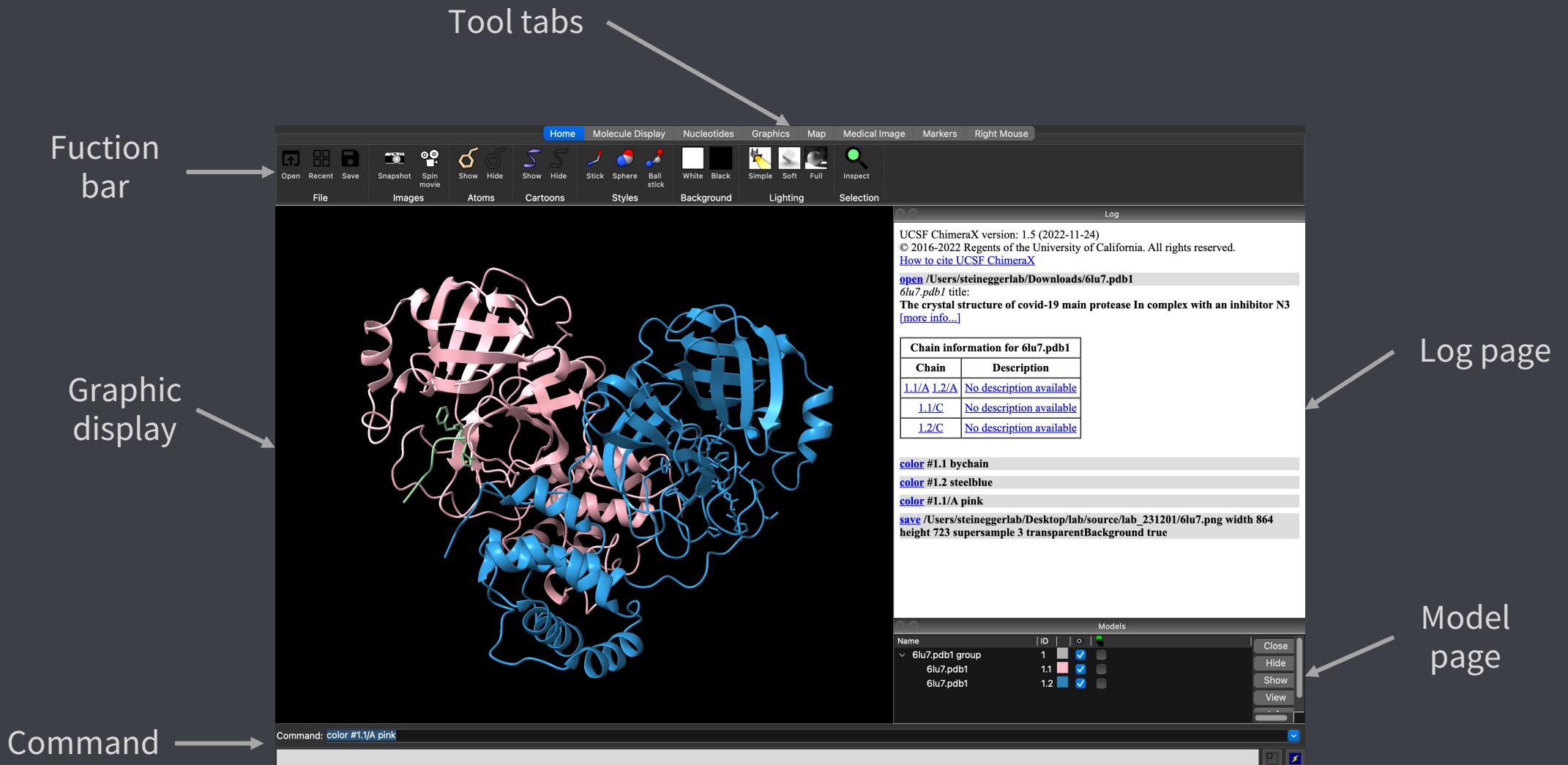


Total 16 Output Files

1 msa file  
1 .done.txt file  
3 plots (pae, coverage, plddt)  
1 PAE value file  
5 score files  
5 pdb files



# ChimeraX UI



# Start with ChimeraX

## Specifiers

model	chain	residue	atom
#	/	:	@

```
#1  
#1/B-D,F  
:12-25,48@ca,n  
:start-40
```

You can even fetch from Online sources  
PDB, AlphaFold, ESMFold, UniProt, etc.

```
open 6lu7  
open p29474 from alphafold
```

## Open, Show, Hide, Close

```
open ( filename | URL | [ prefix:]identifier )  
close model-spec  
show spec [ level | target string ]  
hide spec [ level | target string ]
```

```
open P0DTD1_rank_001.pdb  
show #1/a surface  
hide surface  
close #1
```

## AlphaFold

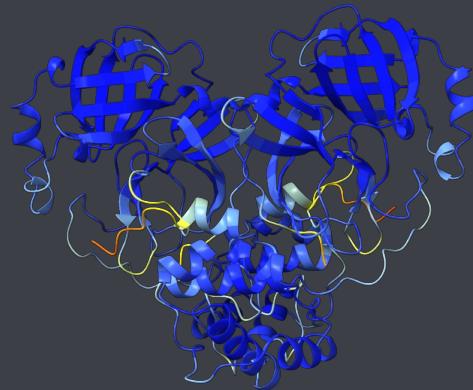
```
alphafold fetch uniprot-id  
alphafold match sequence  
alphafold predict sequence
```

Fetch from AFDB  
Models with similar seq.  
Predict structure

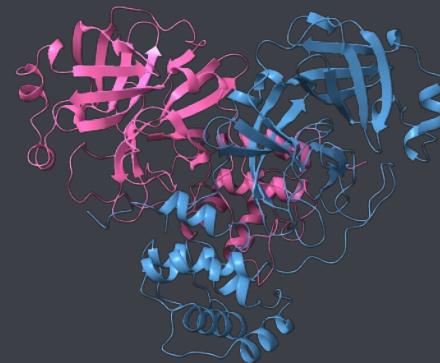
# Molecular Display in ChimeraX

## Color

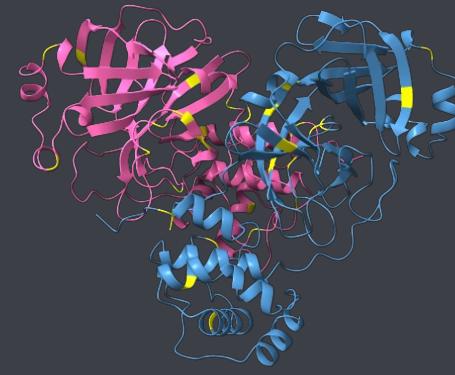
color bfactor palette alphafold



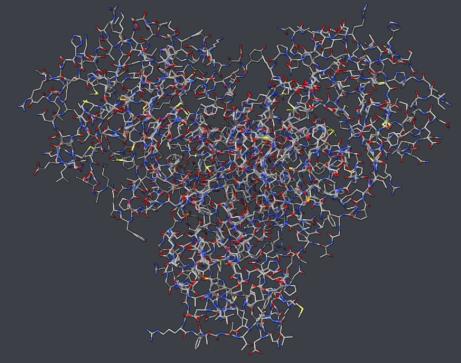
color #1/A hotpink  
color #1/B steelblue



color :ser yellow

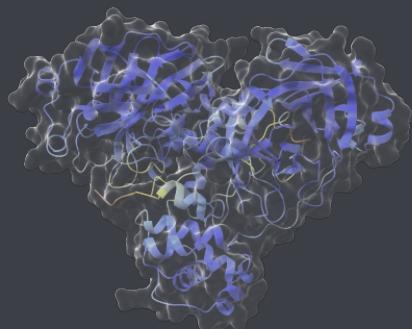


hide cartoon; show atom  
color byatom

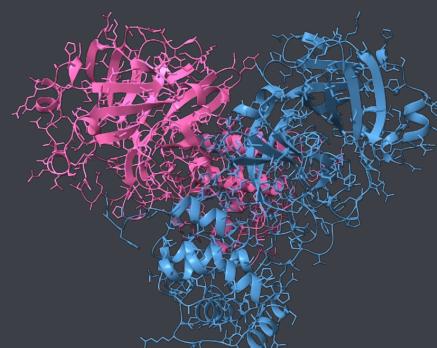


## Surface, Cartoon,Atom, Style

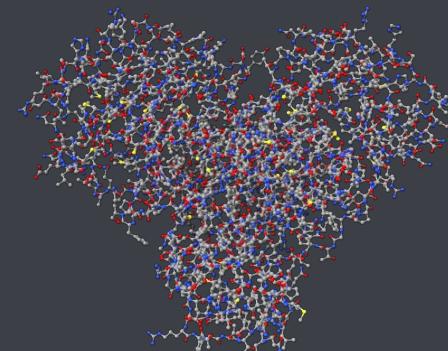
surface color white transparency 60



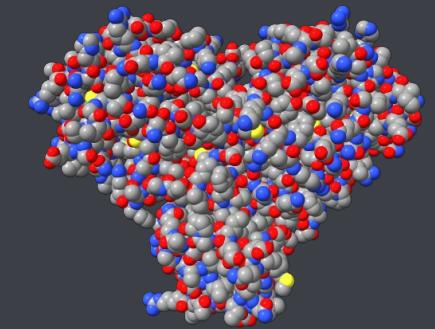
show atoms  
style stick



style ball

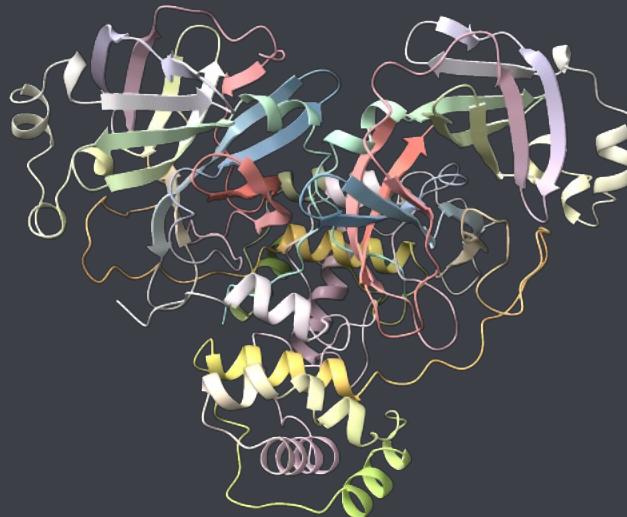


style sphere



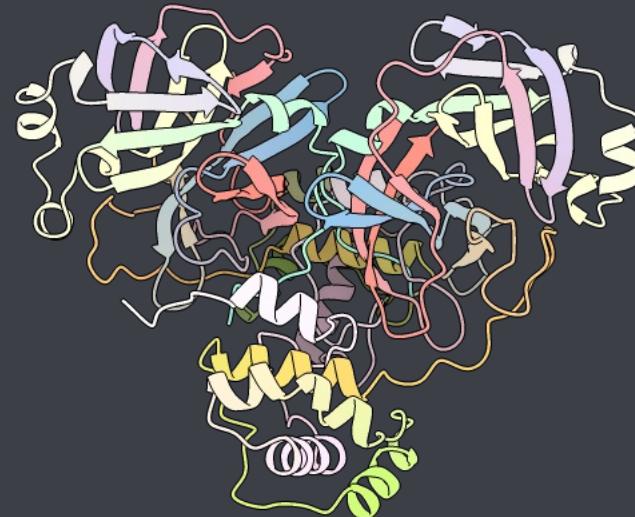
# Graphics in ChimeraX

rainbow palette Set3-9  
lighting simple



lighting

lighting flat



Graphics

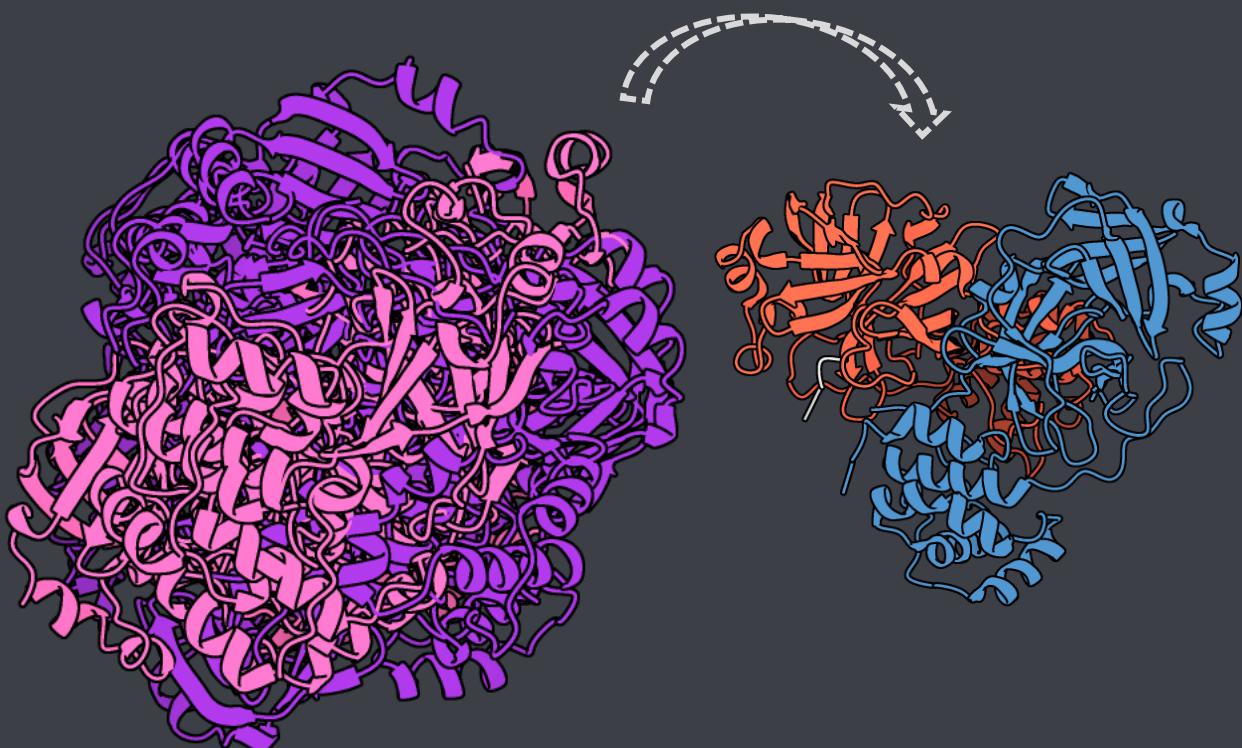
graphics silhouette width 2.5



# Superpose in ChimeraX

Load all 5 models + 6lu7

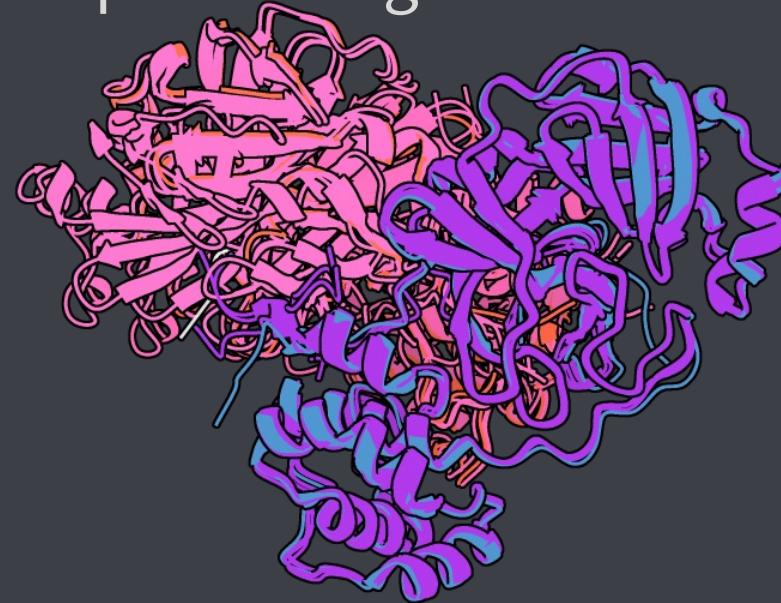
```
open *.pdb  
color /A hotpink  
color /B dark orchid  
color #6/tomato; color #6.2 steelblue
```



Match maker

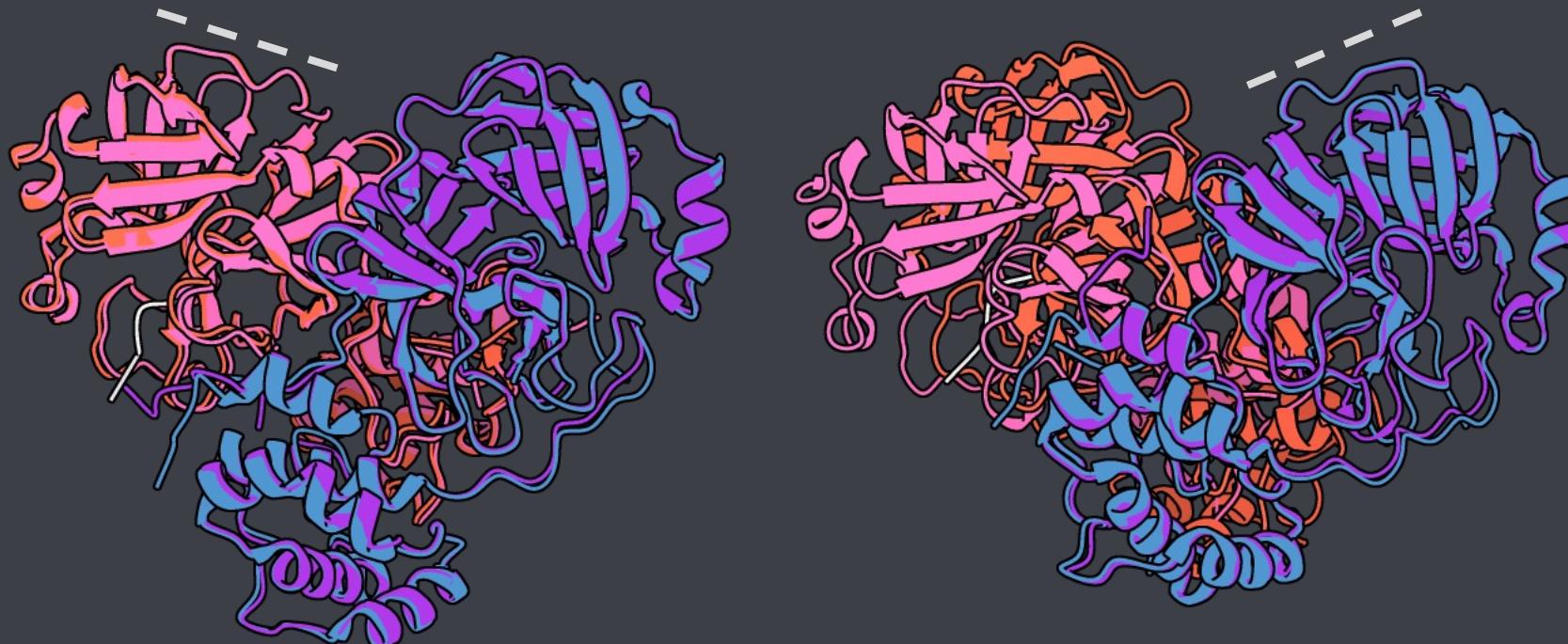
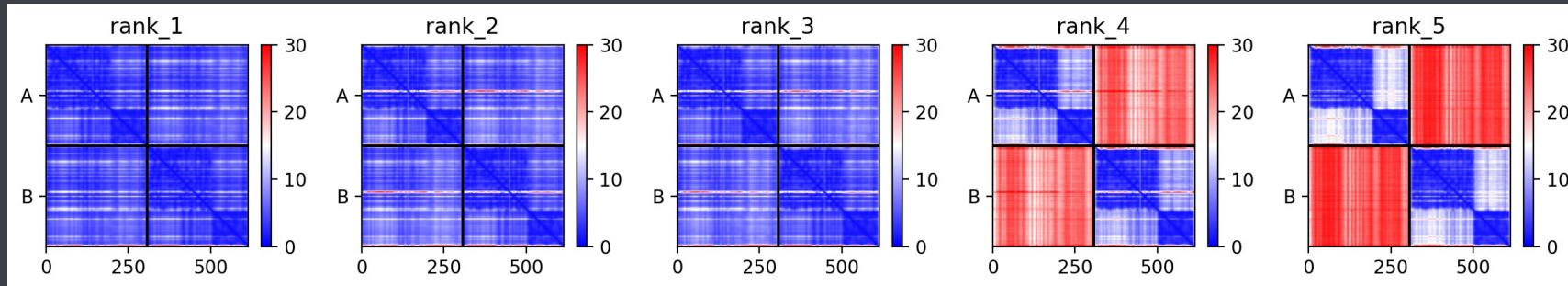
```
( matchmaker | mmaker ) matchstruct to refstruct  
[ bring other-models ] options seq-align-scoring
```

Sequence Alignment based



To superpose by **structure**, use “**align**”  
Caution: Number of atoms should be exactly matched!

# Superpose in ChimeraX



Let's check the white box!

# Tips

## COLOR NAMES

<https://www.cgl.ucsf.edu/chimerax/docs/user/commands/colornames.html>

## SCRIPT(.cxc)

Write all commands in one file and open script file

## Name & Alias

```
alias ribcolor color $1 $2 targ c  
ribcolor '/a & helix' 'dodger blue'
```

```
name tm1 /a:34-64
```

```
color tm1 medium blue
```

## Python script

Define your own function with python!

<https://rbvi.github.io/chimerax-recipes/>  
<https://www.rbvi.ucsf.edu/trac/chimera/wiki/Scripts#ChimeraXPythonScripts>

# All-in-one Notebook

AlphaFold Prediction Analysis Notebook

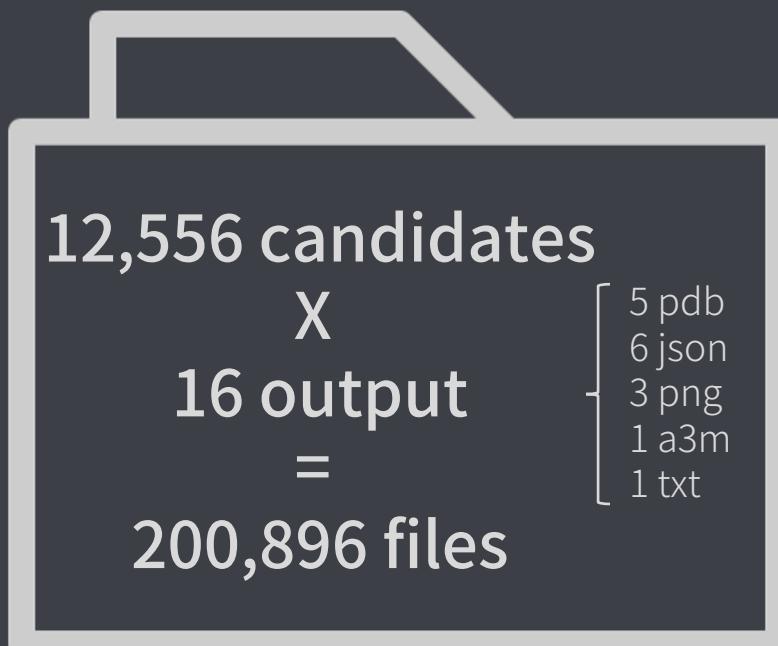
# Why?

Disordered  
Protein  
Interaction



Massive  
Data

Massive  
managers



# Why? During the meetings…



- Present **slides**
- Copy** data from remote to local
- Execute **ChimeraX**
- Enter the commands
- Open image files
- Search through **websites**



How's the structure of model X look like?

Where is the interfaces?

Hmm, what about pLDDT?

Does their PAE looks good?

What is their description?

Rank4 looks interesting..  
How's the structure look like?

# Why? During the meetings…



P0DTD1\_P0DTD1.a3m  
P0DTD1\_P0DTD1.done.txt  
P0DTD1\_P0DTD1\_coverage.png  
P0DTD1\_P0DTD1\_pae.png  
P0DTD1\_P0DTD1\_plddt.png  
P0DTD1\_P0DTD1\_predicted\_aligned\_error\_v1.json  
P0DTD1\_P0DTD1\_scores\_rank\_001\_alphafold2\_multimer\_v3\_model\_5\_seed\_000.json  
P0DTD1\_P0DTD1\_unrelaxed\_rank\_001\_alphafold2\_multimer\_v3\_model\_5\_seed\_000.pdb

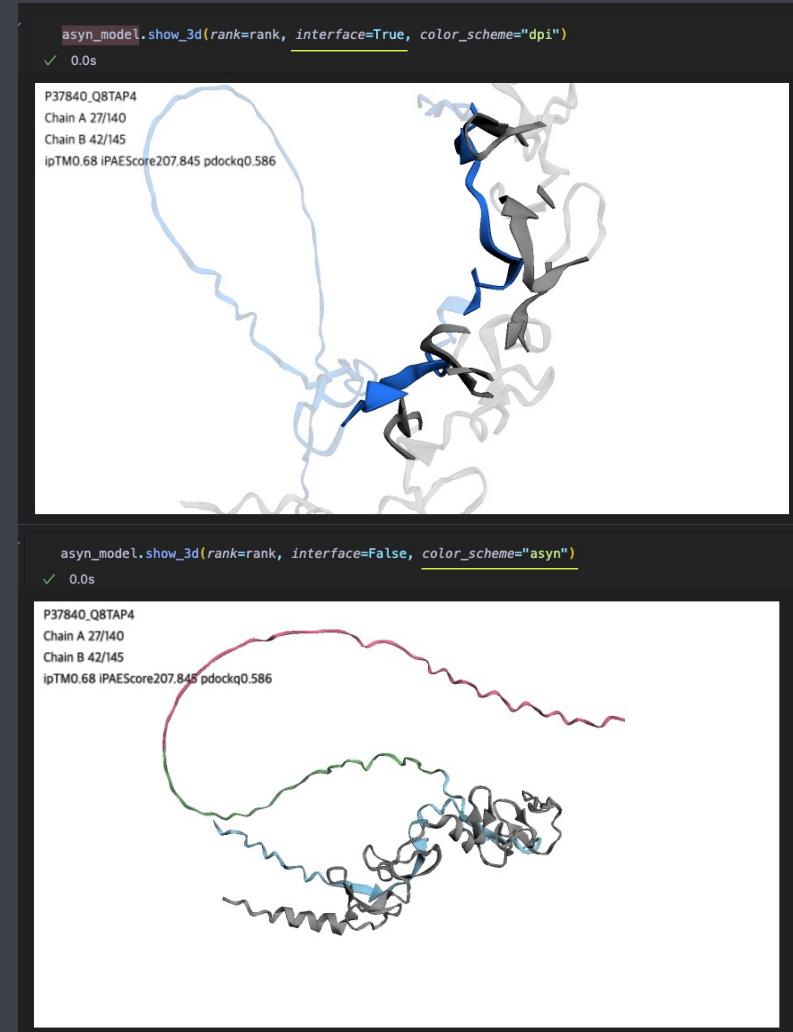
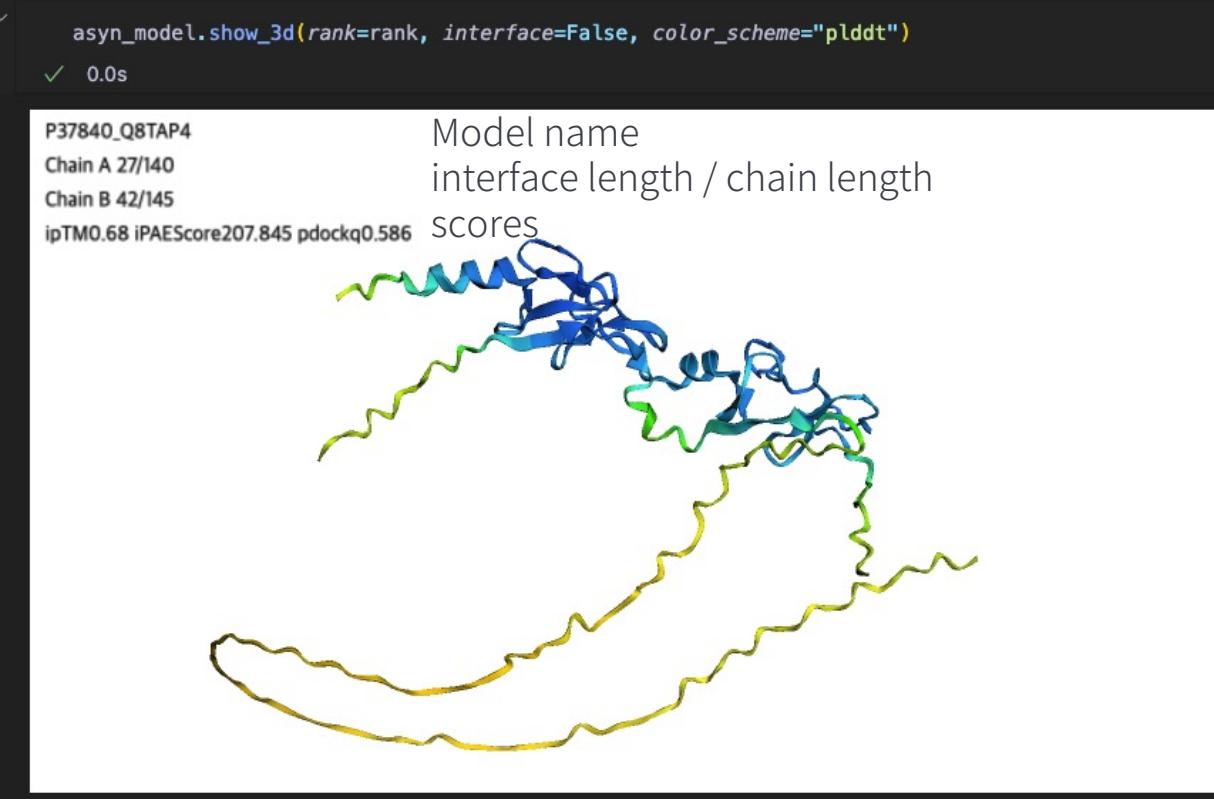
However, these files and questions have **patterns!**

Let's get rid of the redundancy

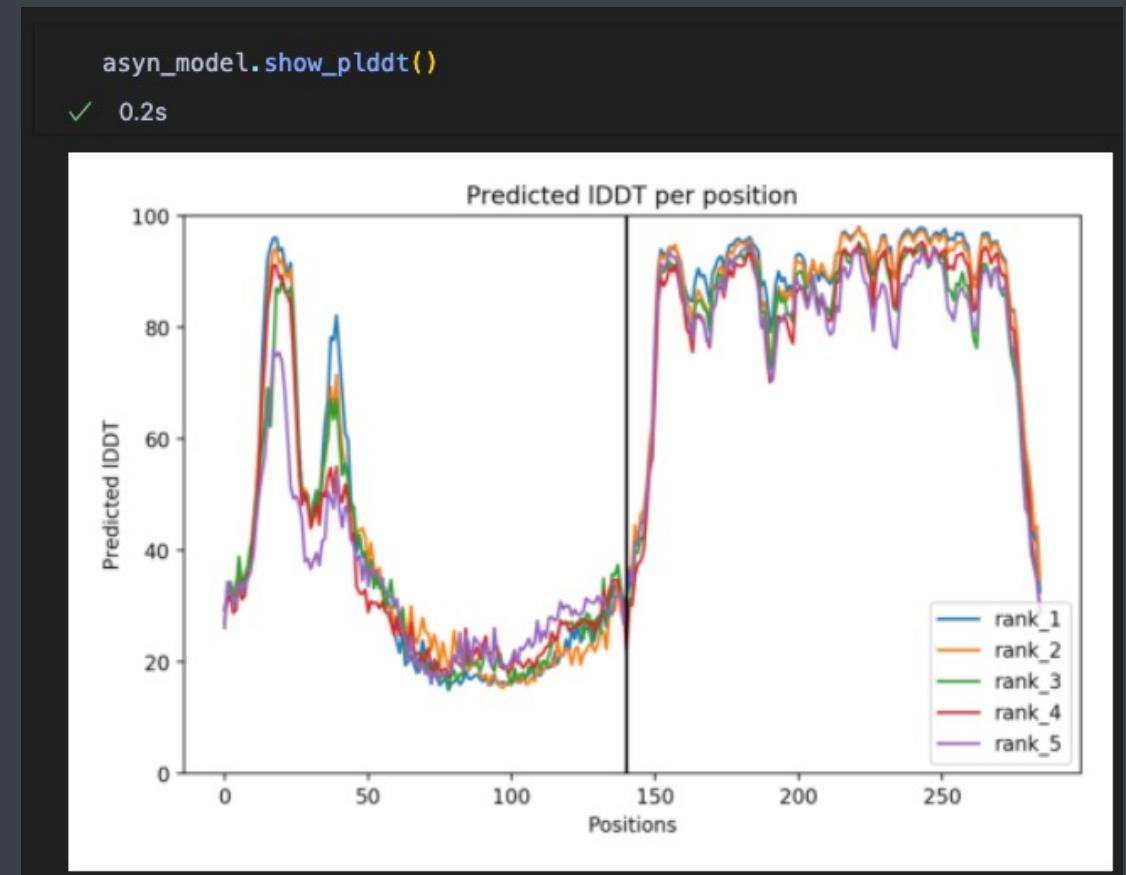
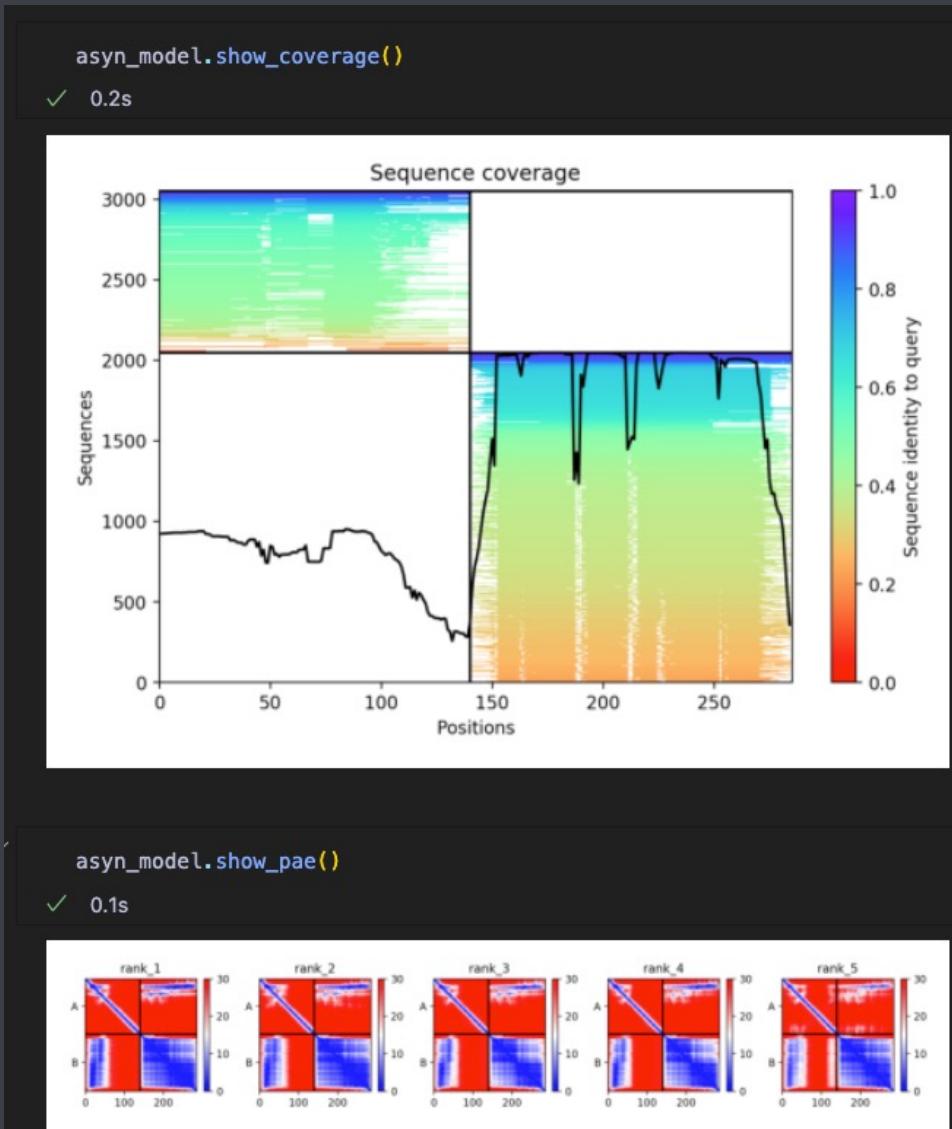
# What? All-in-one Notebook

## Show prediction outputs

- Show 3D structure : interface(True, False), color scheme(dpi, asyn, chain, ss, plddt, plddt\_dpi)
- show plots : pLDDT, PAE, coverage



# What? All-in-one Notebook



# What? All-in-one Notebook

## Get UniProt information

```
| methods: get_info(), get_description(), get_comment(), get_goterms(), get_gene()
```

```
info = asyn_model.get_info()
print(info.keys())
✓ 15.6s

Getting information of Q8TAP4
Getting gene name of Q8TAP4
Getting description of Q8TAP4
Getting GO terms of Q8TAP4
Getting comment of Q8TAP4
dict_keys(['gene', 'description', 'GO', 'INTERACTION', 'ALTERNATIVE PRODUCTS'])
```

```
print(info['gene'])
print(info['description'])
print(info['GO']['biological process'])
print(info['INTERACTION'])
✓ 0.0s

['LMQ3']
['LIM domain only protein 3', 'Neuronal-specific transcription factor DAT1', 'Rhombotin-3']
['acute-phase response [GO:0006953]', 'inflammatory response [GO:0006954]', 'maintenance of gastrointestinal epithelium [GO:0030277]', 'regulation of lipid metabolic process [GO:0019216]']
Q8TAP4; Q8N9N5: BANP; NbExp=3; IntAct=EBI-742259, EBI-744695; Q8TAP4; Q8NA61: CBY2; NbExp=3; IntAct=EBI-742259, EBI-741724; Q8TAP4; Q8IYX8-2: CEP57L1; NbExp=3; IntAct=EBI-742259, EBI-101819
```

# How?

1. Remote Accessibility
2. Prediction outputs (scores, coverage, etc.)
3. Structure visualization + functionality
4. Information about protein

IDE  
Patterns in file names  
Jupyter Notebook + py3Dmol  
Bio, bioservices library

input  
output

Path\_to\_Directory  
**Prediction**  
Prediction class

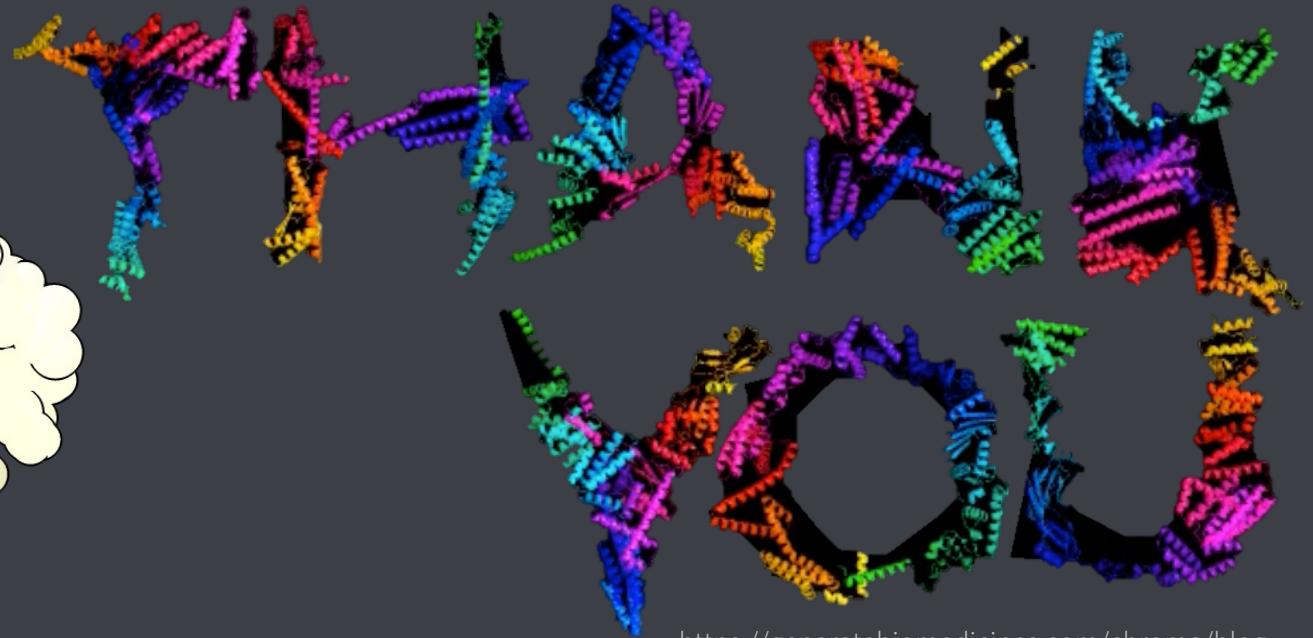
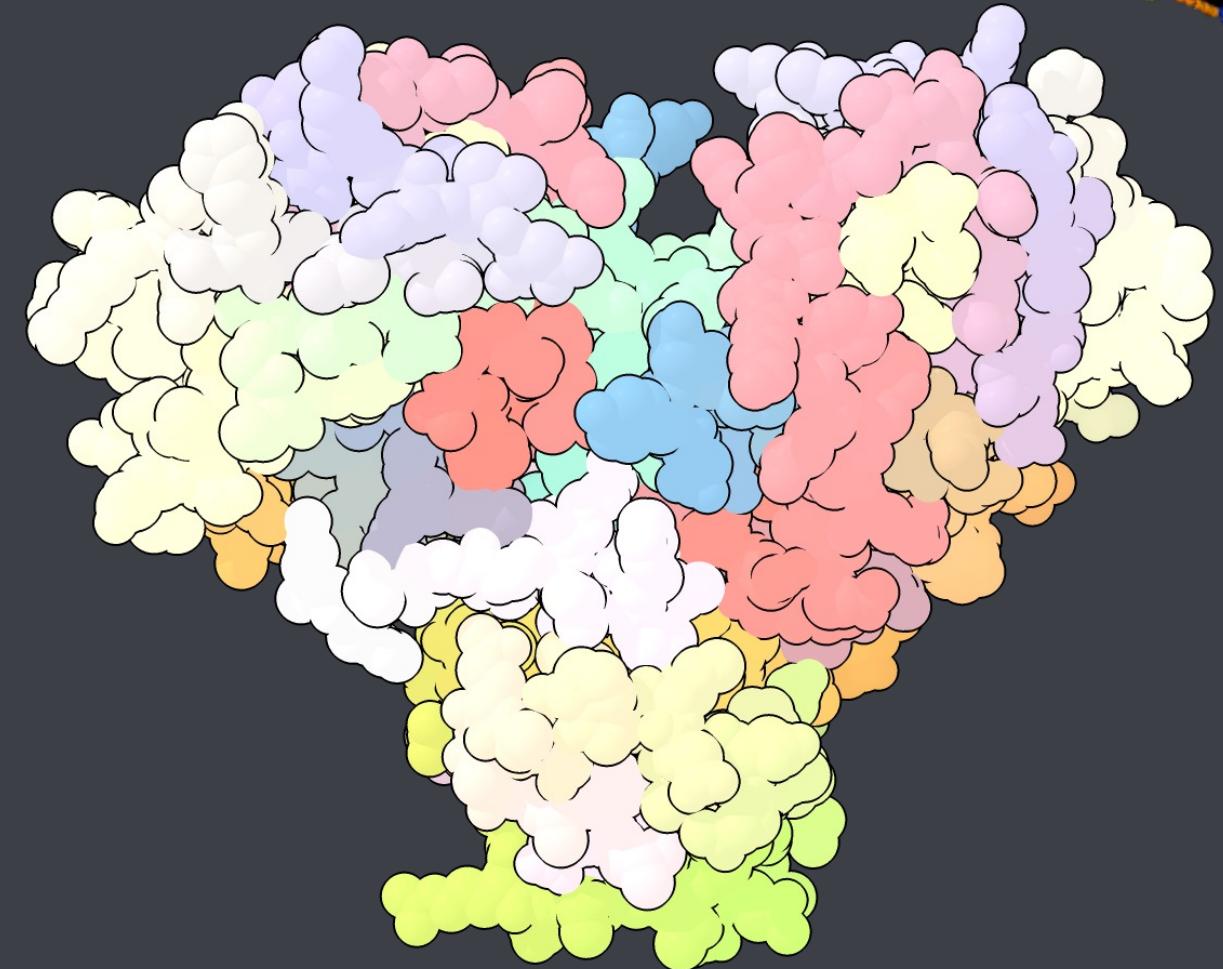
```
● ● ●  
asyn_path = get_path("asyn")  
asyn_pred = Prediction(asyn_path)
```

model\_name, rank  
**Model**  
Model class

```
● ● ●  
#TODO : Define parameters  
partner, rank = "Q8TAP4", 1  
  
asyn_model = asyn_pred.load_model(f"P37840_{partner}")  
asyn_model.get_model(rank = rank)
```

rank, parameters  
**Structure**  
3D structure

```
● ● ●  
asyn_model.show_3d(rank=rank,  
                   interface=False,  
                   color_scheme="asyn")
```



<https://generatebiomedicines.com/chroma/blog>

VISIT HERE!

UCSF ChimeraX ([link](#))

Coda Page ([link](#))

GitHub ([Today's](#))