FoldSeek: Fast search and comparison of protein structures

INTRODUCTION

For today's workshop, we're going to be running through an example of how to create a database of protein structures and query the database for structural similarity to a desired protein. We'll also be going over how to directly compare the structure of two proteins directly against each other, and display the results in tab-delimited and HTML format.

SET UP **[## Estimated runtime: < 2 minutes]**

1. Login to Atlas Open OnDemand using your web browser (https://atlas-ood.hpc.msstate.edu/) and navigate to your working directory for this workshop. Mine, for example, is in the shared directory under my username (Files > /90daydata > Change directory > /90daydata/shared/olivia.haley)

2. Once in your working directory, select **Open in Terminal**. A new window should open.

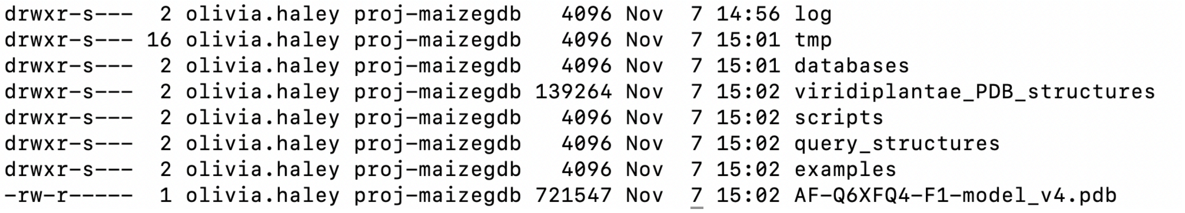
3. Copy the shared directory containing the scripts and structures for this demo to your working directory, then activate the conda environment for this workshop.

```
# Copy the shared directory from the shared folder
cp -r /90daydata/shared/protein_structure_workshop/FoldSeek/ .

# Activate the workshop conda environment
module load miniconda3
source activate /90daydata/shared/protein_structure_conda/foldseek_env

#Navigate into the directory and view its contents
cd FoldSeek
ls -ltr
```

The directory's contents should look like the following:



```
drwxr-s---  2 olivia.haley proj-maizegdb   4096 Nov  7 14:56 log
drwxr-s--- 16 olivia.haley proj-maizegdb   4096 Nov  7 15:01 tmp
drwxr-s---  2 olivia.haley proj-maizegdb   4096 Nov  7 15:01 databases
drwxr-s---  2 olivia.haley proj-maizegdb 139264 Nov  7 15:02 viridiplantae_PDB_structures
drwxr-s---  2 olivia.haley proj-maizegdb   4096 Nov  7 15:02 scripts
drwxr-s---  2 olivia.haley proj-maizegdb   4096 Nov  7 15:02 query_structures
drwxr-s---  2 olivia.haley proj-maizegdb   4096 Nov  7 15:02 examples
-rw-r-----  1 olivia.haley proj-maizegdb 721547 Nov  7 15:02 AF-Q6XFQ4-F1-model_v4.pdb
```

TUTORIAL

The first step is to create the target database, which will contain the structures that you will search against using your query protein(s). Target databases can be established using a directory of protein structure files or fasta files. Some pre-compiled databases are available for downloading directly. In this tutorial, we'll be looking for structural homologs of proteins with potential applications in agriculture, such as:

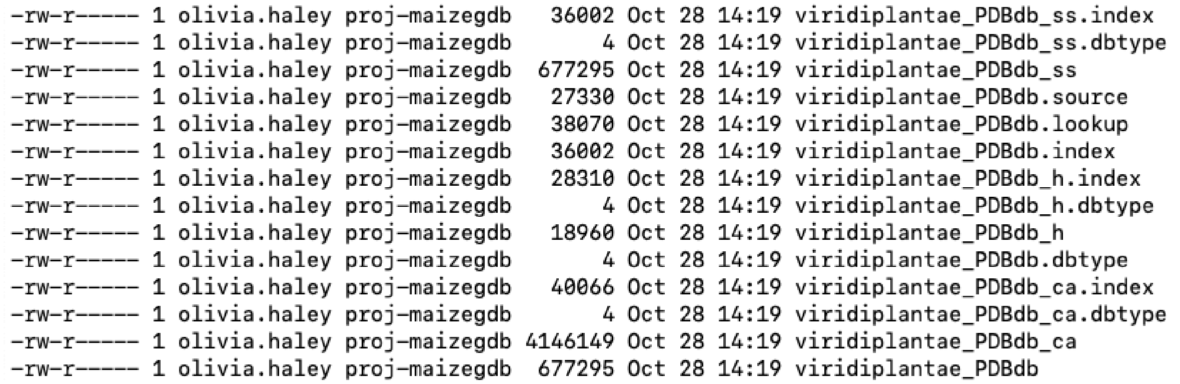
- KWL1, a defense protein in maize with antifungal properties [6FPG]
- GST-I, an enzyme contributing to herbicide detoxification in maize [1AXD]
- PPO, an enzyme which leads to browning in apple [6ELS]

For simplicity, our target database will be a subset of experimentally-determined structures from the *Viridiplantae* (or 'green plants') clade.

Step 1. Create the target database [## Estimated runtime: < 1 minute]

```
#Run the script to create the target database of Viridiplantae structures
sbatch scripts/s0_create_database.sh

#View the files generated while creating the database
ls -ltr databases
```

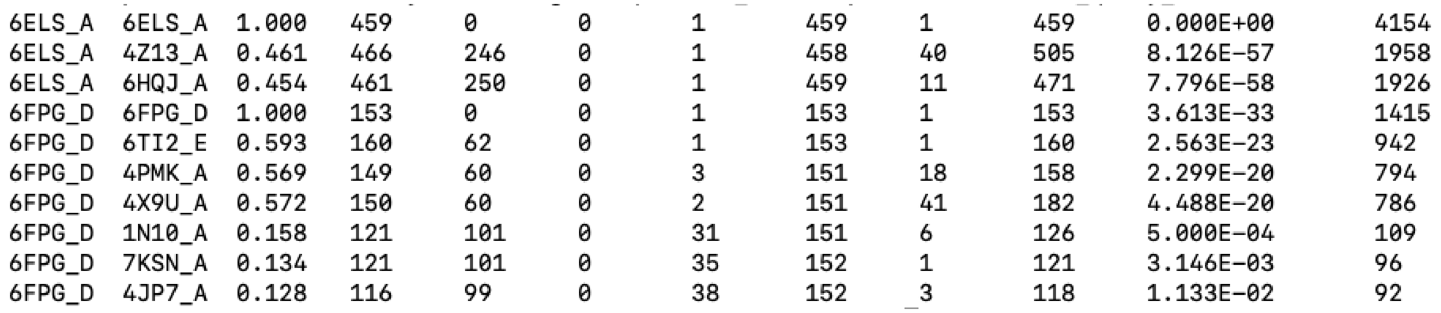


```
-rw-r----- 1 olivia.haley proj-maizegdb   36002 Oct 28 14:19 viridiplantae_PDBdb_ss.index
-rw-r----- 1 olivia.haley proj-maizegdb       4 Oct 28 14:19 viridiplantae_PDBdb_ss.dbtype
-rw-r----- 1 olivia.haley proj-maizegdb  677295 Oct 28 14:19 viridiplantae_PDBdb_ss
-rw-r----- 1 olivia.haley proj-maizegdb   27330 Oct 28 14:19 viridiplantae_PDBdb.source
-rw-r----- 1 olivia.haley proj-maizegdb   38070 Oct 28 14:19 viridiplantae_PDBdb.lookup
-rw-r----- 1 olivia.haley proj-maizegdb   36002 Oct 28 14:19 viridiplantae_PDBdb.index
-rw-r----- 1 olivia.haley proj-maizegdb   28310 Oct 28 14:19 viridiplantae_PDBdb_h.index
-rw-r----- 1 olivia.haley proj-maizegdb       4 Oct 28 14:19 viridiplantae_PDBdb_h.dbtype
-rw-r----- 1 olivia.haley proj-maizegdb   18960 Oct 28 14:19 viridiplantae_PDBdb_h
-rw-r----- 1 olivia.haley proj-maizegdb       4 Oct 28 14:19 viridiplantae_PDBdb.dbtype
-rw-r----- 1 olivia.haley proj-maizegdb   40066 Oct 28 14:19 viridiplantae_PDBdb_ca.index
-rw-r----- 1 olivia.haley proj-maizegdb       4 Oct 28 14:19 viridiplantae_PDBdb_ca.dbtype
-rw-r----- 1 olivia.haley proj-maizegdb 4146149 Oct 28 14:19 viridiplantae_PDBdb_ca
-rw-r----- 1 olivia.haley proj-maizegdb  677295 Oct 28 14:19 viridiplantae_PDBdb
```

Step 2. Run FoldSeek (TSV output) to get initial query matches [## Estimated runtime: < 1 minute]

```
#Run the script
sbatch scripts/s1_foldseek_run_initial_query.sh

#View the output tab-delimited file
head initial_query_results.tsv
```

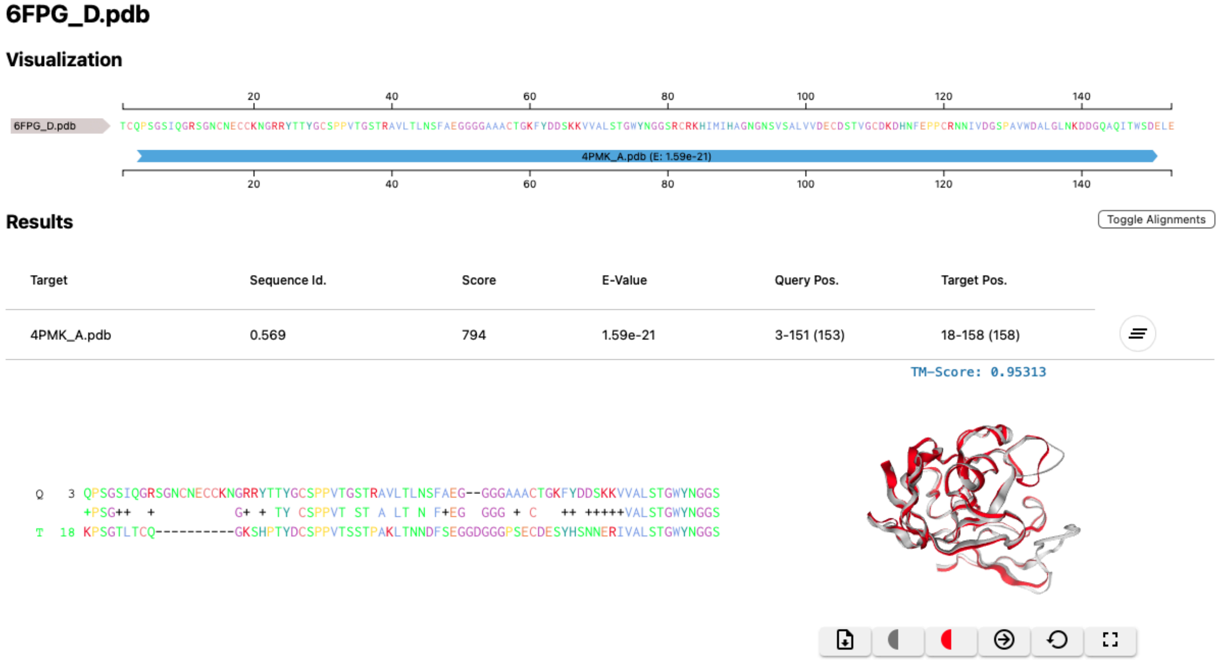


```
6ELS_A  6ELS_A  1.000   459     0       0       1       459     1       459     0.000E+00       4154
6ELS_A  4Z13_A  0.461   466     246     0       1       458     40      505     8.126E-57       1958
6ELS_A  6HQJ_A  0.454   461     250     0       1       459     11      471     7.796E-58       1926
6FPG_D  6FPG_D  1.000   153     0       0       1       153     1       153     3.613E-33       1415
6FPG_D  6TI2_E  0.593   160     62      0       1       153     1       160     2.563E-23       942
6FPG_D  4PMK_A  0.569   149     60      0       3       151     18      158     2.299E-20       794
6FPG_D  4X9U_A  0.572   150     60      0       2       151     41      182     4.488E-20       786
6FPG_D  1N10_A  0.158   121     101     0       31      151     6       126     5.000E-04       109
6FPG_D  7KSN_A  0.134   121     101     0       35      152     1       121     3.146E-03       96
6FPG_D  4JP7_A  0.128   116     99      0       38      152     3       118     1.133E-02       92
```

Step 3. Run FoldSeek (HTML output) to explore the initial query-target matches [## Estimated runtime: < 1 minute]

We have some matches from each of our proteins, but let’s explore one of the matches for maize kiwellin defense protein (6FPG_D). In particular, let’s look at the alignment between the kiwellin and 4PMK_A. This script will output the alignment’s results in HTML form as a file called ***6FPGD_4PMKA.html***. To move the file to your local machine using the Atlas Open On Demand interface, click on the open tab, then Refresh > FoldSeek). Next to the file name there should be a drop-down box, click the drop-down then select Download.

```
#Run the script
sbatch scripts/s2_view_query_target_alignment.sh
```



6FPG_D.pdb

Visualization

20
40
60
80
100
120
140
6FPG_D.pdb
TCQPSGSIQGRSGNCNECCKNGRRYTTYGCSPPVTGSTRAVLTLNSFAEGGGGAAACTGKFYDDSKKVVALSTGWYNGGSRCRKHIMIHAGNGNSVSALVVDECDSTVGCDKDHNFEPPCRNNIVDGSPAVWDALGLNKDDGQAQITWSDELE
4PMK_A.pdb (E: 1.59e-21)
20
40
60
80
100
120
140

Results

Target	Sequence Id.	Score	E-Value	Query Pos.	Target Pos.
4PMK_A.pdb	0.569	794	1.59e-21	3-151 (153)	18-158 (158)

TM-Score: 0.95313

```
Q   3 QPSGSIQGRSGNCNECCKNGRRYTTYGCSPPVTGSTRAVLTLNSFAEG--GGGAAACTGKFYDDSKKVVALSTGWYNGGS
      +PSG++  +          G+ + TY CSPPVT ST A LT N F+EG  GGG + C   ++ +++++VALSTGWYNGGS
T  18 KPSGTLTCQ----------GKSHPTYDCSPPVTSSTPAKLTNNDFSEGGDGGGPSECDESYHSNNERIVALSTGWYNGGS
```

As it turns out, this protein 4PMK_A is also a kiwellin! It comes from *Actinidia chinensis* var. *chinensis* (the Chinese soft-hair kiwi) where it was first identified as an allergen (Tamburrini et al., 2005). Its homolog in maize was later found to have antifungal properties (Han et al., 2019).

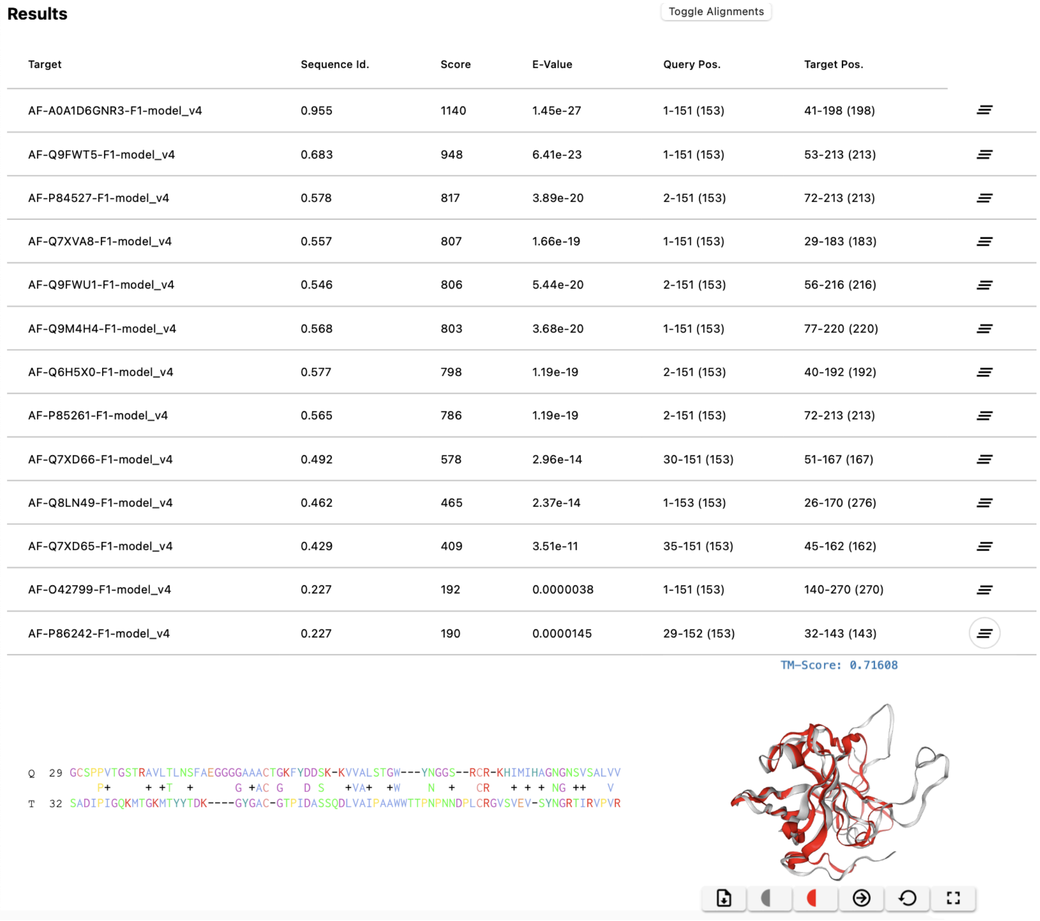
Bonus Question: What insights can you make from the other proteins?

Step 4. Expand the search to other clades [## Estimated time: < 2 minutes]

Until this point, we’ve been using a database of experimentally-determined protein structures. But less than 10% of the structures in the Protein Data Bank come from plants! For this exercise, we'll expand our structural homology search by using a database of computational protein structures in FoldSeek. FoldSeek has pre-compiled databases such as AlphaFoldDB, PDB, and ESMAtlas. In this example we’ll use the AlphaFold database of Swiss-Prot proteins. This script will create an HTML file that you’ll need to transfer to your local computer with Atlas Open On Demand.

```
#Generates the database files for the AlphaFold/Swiss-Prot database
#Runs the query protein against the AlphaFold/Swiss-Prot database
sbatch scripts/s3_create_and_query_AF2_database.sh
```

Looking at the results, many of the top hits are kiwellins in plant species like rice (*Oryzae sativa*) and kiwi. What's interesting is that we do see structural homology for our maize kiwellin and a protein from *Streptomyces mobaraensis*, a spore-forming bacterium that is known to produce antimicrobial compounds (Zindel et al., 2013; P86242).



Results

Target	Sequence Id.	Score	E-Value	Query Pos.	Target Pos.
AF-A0A1D6GNR3-F1-model_v4	0.955	1140	1.45e-27	1-151 (153)	41-198 (198)
AF-Q9FWT5-F1-model_v4	0.683	948	6.41e-23	1-151 (153)	53-213 (213)
AF-P84527-F1-model_v4	0.578	817	3.89e-20	2-151 (153)	72-213 (213)
AF-Q7XVA8-F1-model_v4	0.557	807	1.66e-19	1-151 (153)	29-183 (183)
AF-Q9FWU1-F1-model_v4	0.546	806	5.44e-20	2-151 (153)	56-216 (216)
AF-Q9M4H4-F1-model_v4	0.568	803	3.68e-20	1-151 (153)	77-220 (220)
AF-Q6H5X0-F1-model_v4	0.577	798	1.19e-19	2-151 (153)	40-192 (192)
AF-P85261-F1-model_v4	0.565	786	1.19e-19	2-151 (153)	72-213 (213)
AF-Q7XD66-F1-model_v4	0.492	578	2.96e-14	30-151 (153)	51-167 (167)
AF-Q8LN49-F1-model_v4	0.462	465	2.37e-14	1-153 (153)	26-170 (276)
AF-Q7XD65-F1-model_v4	0.429	409	3.51e-11	35-151 (153)	45-162 (162)
AF-O42799-F1-model_v4	0.227	192	0.0000038	1-151 (153)	140-270 (270)
AF-P86242-F1-model_v4	0.227	190	0.0000145	29-152 (153)	32-143 (143)

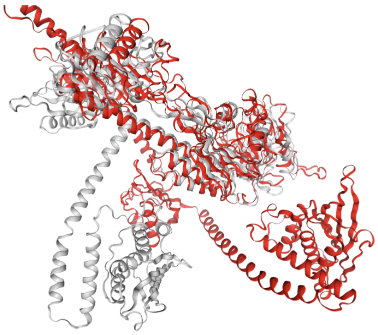
TM-Score: 0.71608

```
Q  29 GCSPPVTGSTRAVLTLNSFAEGGGGAAACTGKFYDDSK-KVVALSTGW---YNGGS--RCR-KHIMIHAGNGNSVSALVV
          P+     + +T  +      G +AC G   D S   +VA+  +W    N  +   CR   + + + NG ++   V
T  32 SADIPIGQKMTGKMTYYTDK----GYGAC-GTPIDASSQDLVAIPAAWWTTPNPNNDPLCRGVSVEV-SYNGRTIRVPVR
```

Step 4. When might FoldSeek not perform as expected? [## Estimated time: < 2 minute]

FoldSeek performs a rigid structural alignment, meaning that it doesn't account for the flexibility of protein backbones during the structural similarity search. There are cases (particularly when using computational protein structures) where this can lead to inaccurate conclusions. For example, let's compare the experimental structure of a maize photosystem I protein, with its AlphaFold2 structure. Run the script, and then download the HTML output file (maize-phytochrome-comparison.html)

```
#Generates the database files for the AlphaFold/Swiss-Prot database
#Runs the query protein against the AlphaFold/Swiss-Prot database
sbatch scripts/s4_compare_protein_structures.sh
```



Although these two are the same protein, this is not a great alignment. The TM-score is < 0.50 (indicating they're not assuming the same general fold). In this case, it looks like we have a protein domain in our computational structure (red) that is not in the same orientation as the domain in the experimental structure (gray).

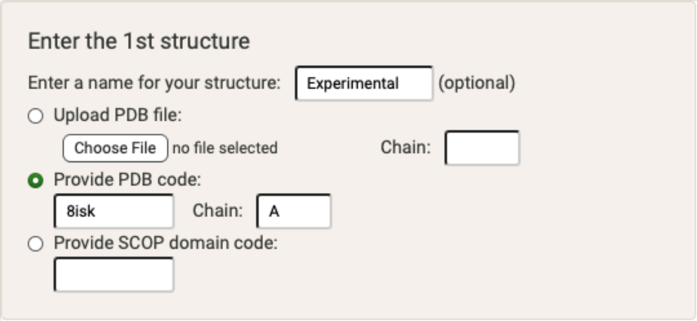
Bonus Question: What can we use to perform the structural alignment and 'fix' the domain orientation?

Step 5 (Optional). Flexible structure alignment [## Estimated time: < 5 minute]

Often, these cases can be corrected by using a flexible structural alignment program. We'll use one such program, called FATCAT to perform the flexible alignment. For a small set of alignments, it'll likely be easier to use their web interface https://fatcat.godziklab.org/fatcat/fatcat_pair.html. However, FATCAT does have a local implementation (https://github.com/GodzikLab/FATCAT-dist) if you have a larger set of structural alignments to perform. On the web platform, you can use PDB codes to upload structure files directly from the PDB, or input your own .pdb files. Note that this may not perform as intended for mmCIF files, or for structures in the PDB which only have mmCIF files available.

Under ' *Enter the 1st structure* ' provide the PDB code **8ISK** and input **A** under Chain. Alternatively, you can download the structure from our workshop directory and upload it. Make sure to select ' *Upload PDB file:* '

Under ' *Enter the 2nd structure* ', select ' *Upload PDB file:* ' and upload the AlphaFold2 structure file of Q6XFQ4 (AF-Q6XFQ4-F1-v4.pdb)



Enter the 1st structure

Enter a name for your structure: Experimental (optional)

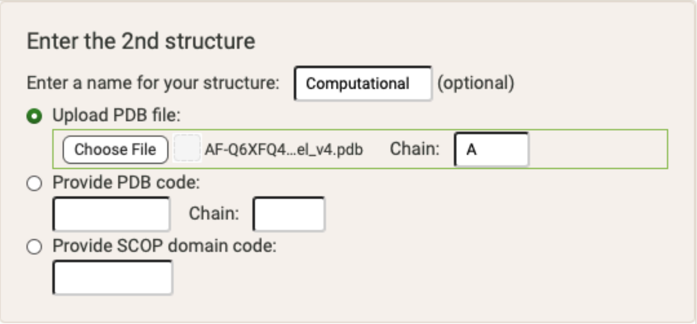
○ Upload PDB file:

Choose File no file selected Chain:

● Provide PDB code:

8isk Chain: A

○ Provide SCOP domain code:



Enter the 2nd structure

Enter a name for your structure: Computational (optional)

● Upload PDB file:

Choose File AF-Q6XFQ4...el_v4.pdb Chain: A

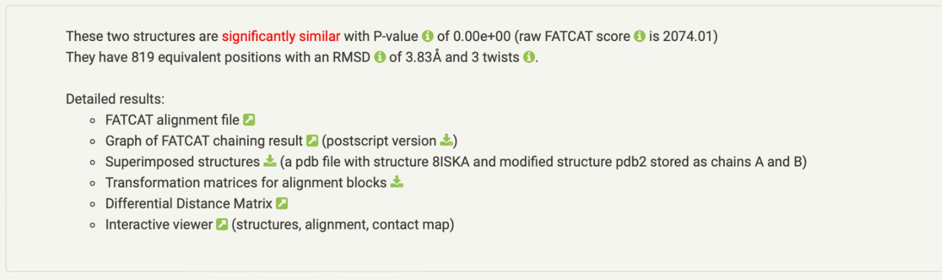
○ Provide PDB code:

Chain:

○ Provide SCOP domain code:

FATCAT will provide a couple of outputs. Of note is the *P*-value which indicates the statistical similarity of the structural alignment (testing the hypothesis if the alignment score occurred randomly). It will also give a breakdown of the number of residues included in the alignment, the RMSD, and how many 'twists' were needed to align the structures. To view the alignment, download the superimposed structures (.pdb file), and then drag the file into the Mol* viewer (https://molstar.org/viewer/).

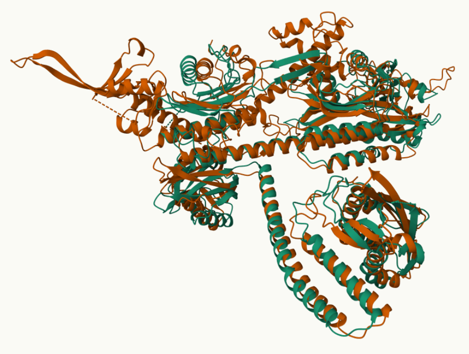
The alignment of these two proteins is much better once we allow for flexibility in the protein backbone!



These two structures are significantly similar with P-value of 0.00e+00 (raw FATCAT score is 2074.01)
They have 819 equivalent positions with an RMSD of 3.83Å and 3 twists .

Detailed results:

- FATCAT alignment file
- Graph of FATCAT chaining result (postscript version)
- Superimposed structures (a pdb file with structure 8ISKA and modified structure pdb2 stored as chains A and B)
- Transformation matrices for alignment blocks
- Differential Distance Matrix
- Interactive viewer (structures, alignment, contact map)



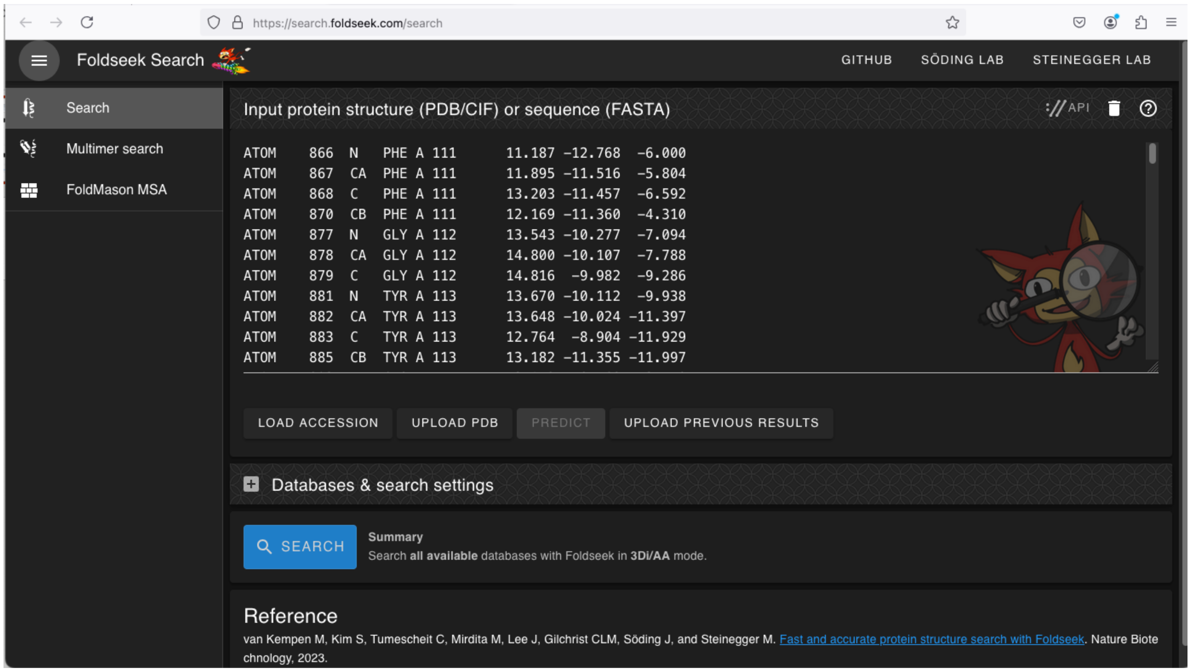
Step 6. Deactivate the environment

```
conda deactivate
```

Foldseek (Online Server)

Foldseek Search online server can be accessed here: https://search.foldseek.com/search

Following the link opens to the search portal.



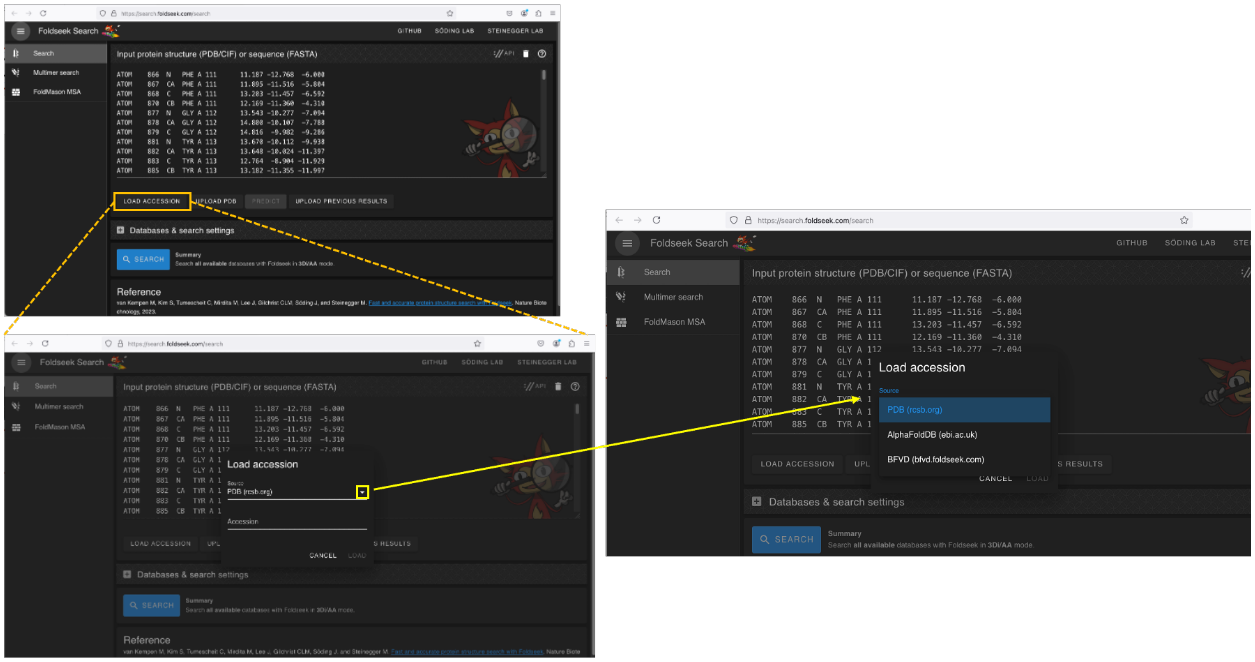
Input protein structure (PDB/CIF) or sequence (FASTA)

```
ATOM    866  N   PHE A 111      11.187 -12.768  -6.000
ATOM    867  CA  PHE A 111      11.895 -11.516  -5.804
ATOM    868  C   PHE A 111      13.203 -11.457  -6.592
ATOM    870  CB  PHE A 111      12.169 -11.360  -4.310
ATOM    877  N   GLY A 112      13.543 -10.277  -7.094
ATOM    878  CA  GLY A 112      14.800 -10.107  -7.788
ATOM    879  C   GLY A 112      14.816  -9.982  -9.286
ATOM    881  N   TYR A 113      13.670 -10.112  -9.938
ATOM    882  CA  TYR A 113      13.648 -10.024 -11.397
ATOM    883  C   TYR A 113      12.764  -8.904 -11.929
ATOM    885  CB  TYR A 113      13.182 -11.355 -11.997
```

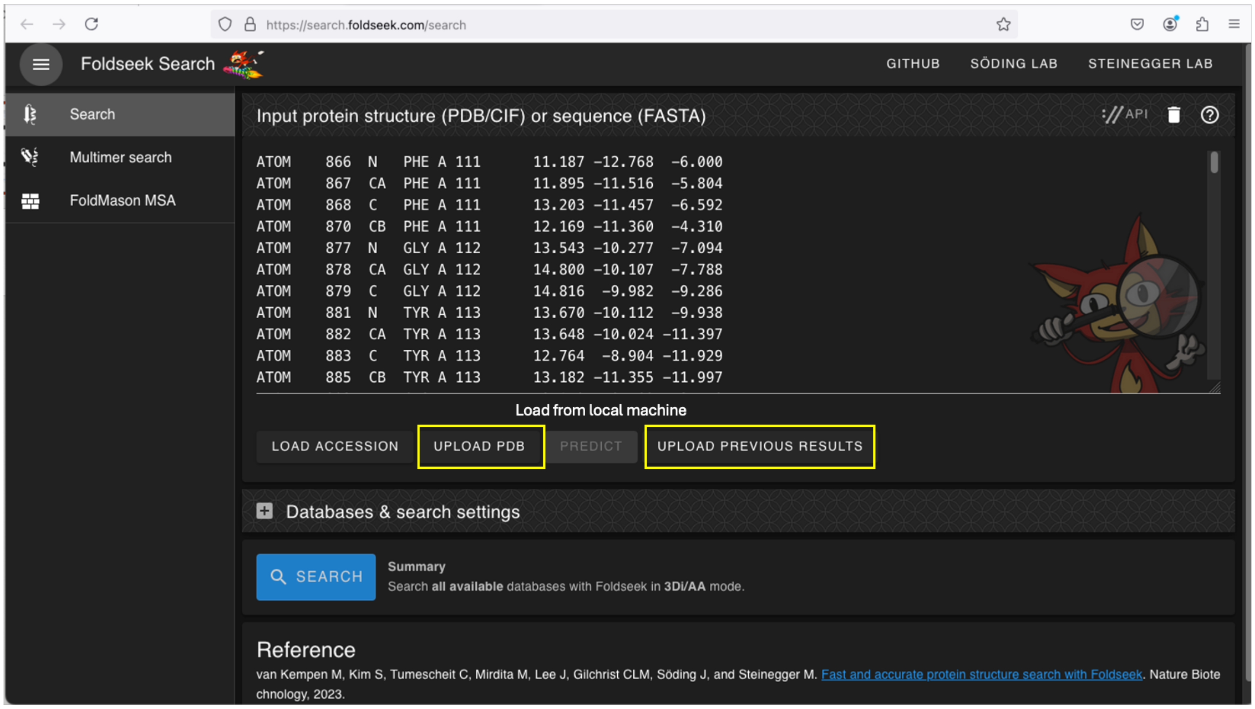
Summary
Search **all available** databases with Foldseek in **3Di/AA** mode.

Reference
van Kempen M, Kim S, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J, and Steinegger M. Fast and accurate protein structure search with Foldseek. Nature Biotechnology, 2023.

Data can be uploaded from a local machine in either PDB or CIF formats using the “UPLOAD PDB” button. The “LOAD ACCESSION” button allows you to import a PDB from a repository (e.g, RCSB Protein Data Bank).



Results from a previous analysis can be uploaded using the “UPLOAD PREVIOUS RESULTS” button.

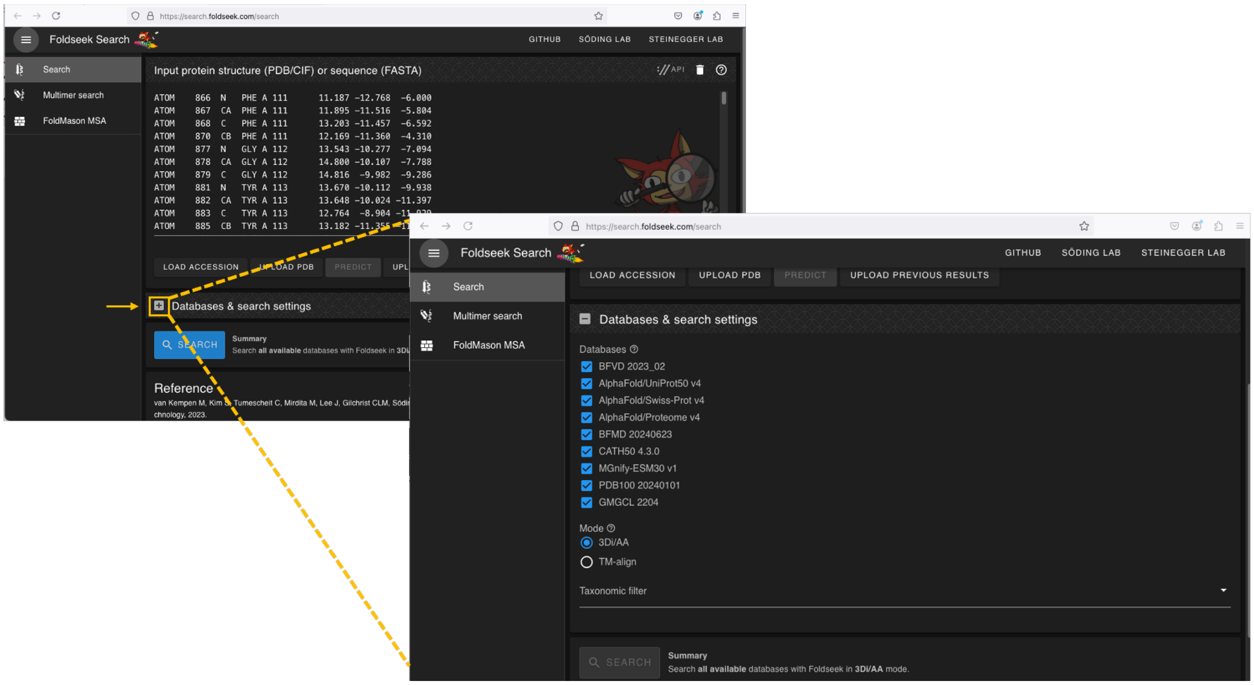


```
ATOM    866  N   PHE A 111      11.187 -12.768  -6.000
ATOM    867  CA  PHE A 111      11.895 -11.516  -5.804
ATOM    868  C   PHE A 111      13.203 -11.457  -6.592
ATOM    870  CB  PHE A 111      12.169 -11.360  -4.310
ATOM    877  N   GLY A 112      13.543 -10.277  -7.094
ATOM    878  CA  GLY A 112      14.800 -10.107  -7.788
ATOM    879  C   GLY A 112      14.816  -9.982  -9.286
ATOM    881  N   TYR A 113      13.670 -10.112  -9.938
ATOM    882  CA  TYR A 113      13.648 -10.024 -11.397
ATOM    883  C   TYR A 113      12.764  -8.904 -11.929
ATOM    885  CB  TYR A 113      13.182 -11.355 -11.997
```

Summary
Search **all available** databases with Foldseek in **3Di/AA** mode.

Reference
van Kempen M, Kim S, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J, and Steinegger M. Fast and accurate protein structure search with Foldseek. Nature Biotechnology, 2023.

If desired, the databases and search settings can be modified (e.g., alignment and taxonomic filtering).



Databases
☑ BFVD 2023_02
☑ AlphaFold/UniProt50 v4
☑ AlphaFold/Swiss-Prot v4
☑ AlphaFold/Proteome v4
☑ BFMD 20240623
☑ CATH50 4.3.0
☑ MGnify-ESM30 v1
☑ PDB100 20240101
☑ GMGCL 2204

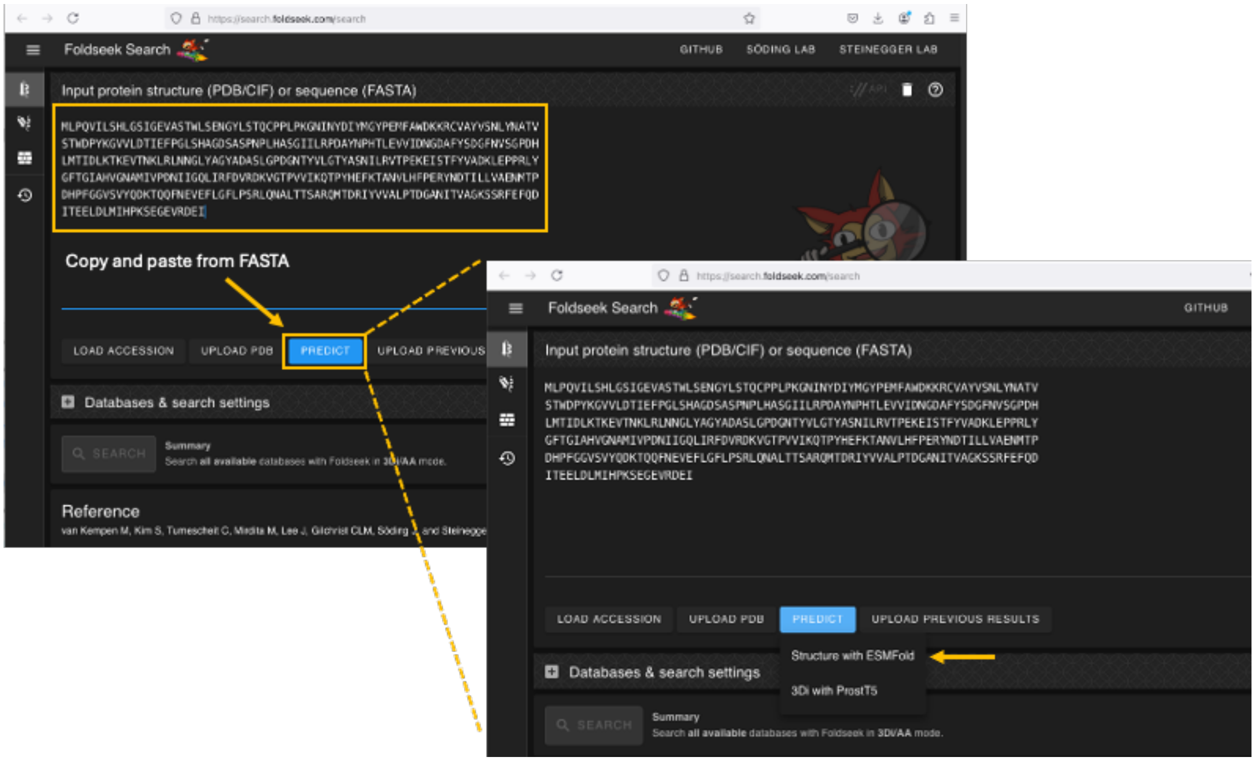
Mode
● 3Di/AA
○ TM-align

Taxonomic filter

Summary
Search **all available** databases with Foldseek in **3Di/AA** mode.

Here we are going to open our FASTA file and past our protein sequence from a FASTA file directly into the input field, click predict, and use ESMFold.

NOTE: FASTA input cannot be uploaded and must be pasted in



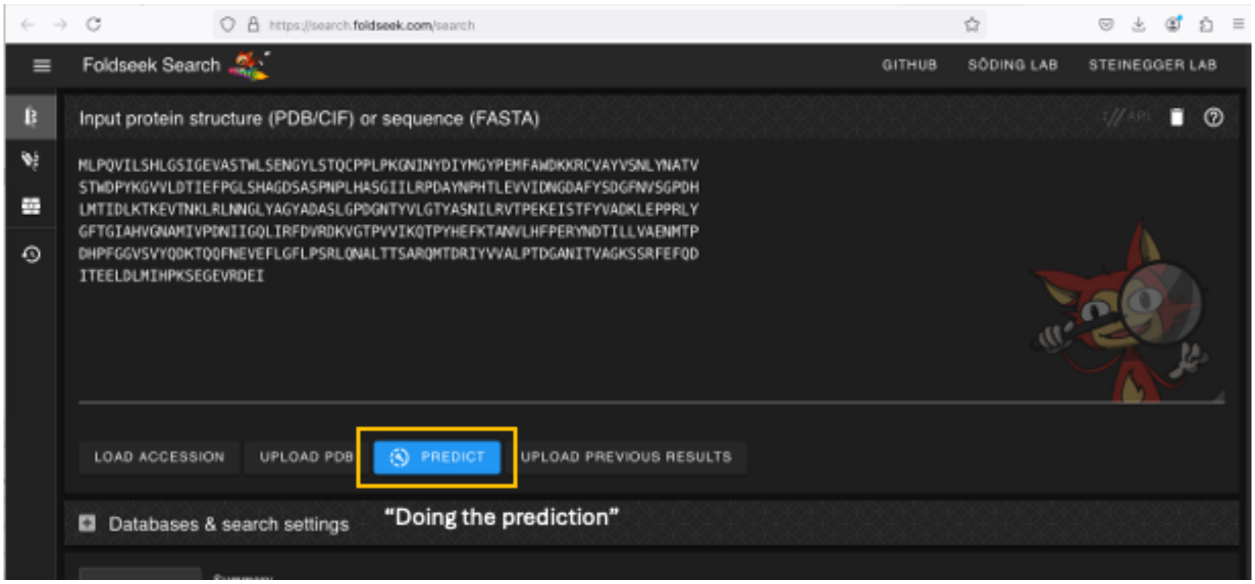
https://search.foldseek.com/search
Foldseek Search
GITHUB
SÖDING LAB
STEINEGGER LAB
Input protein structure (PDB/CIF) or sequence (FASTA)

```
MLPQVILSHLGSIGEVASTWLSENGYLSTQCPPLPKGNINYDIYMGYPEMFAWDKKRCVAYVSNLYNATV
STWDPYKGVVLDTIEFPGLSHAGDSASPNPLHASGIILRPDAYNPHTLEVVIDNGDAFYSDGFNVSGPDH
LMTIDLKTKEVTNKLRLNNGLYAGYADASLGPDGNTYVLGTYASNILRVTPEKEISTFYVADKLEPPRLY
GFTGIAHVGNAMIVPDNIIGQLIRFDVRDKVGTPVVIKQTPYHEFKTANVLHFPERYNDTILLVAENMTP
DHPFGGVSVYQDKTQQFNEVEFLGFLPSRLQNALTTSARQMTDRIYVVALPTDGANITVAGKSSRFEFQD
ITEELDLMIHPKSEGEVRDEI
```

Copy and paste from FASTA

LOAD ACCESSION
UPLOAD PDB
PREDICT
UPLOAD PREVIOUS RESULTS
Structure with ESMFold
3Di with ProstT5
Databases & search settings
SEARCH
Summary
Search all available databases with Foldseek in 3Di/AA mode.
Reference
van Kempen M, Kim S, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J, and Steinegger

After selecting the method, generate the prediction by clicking “predict”.

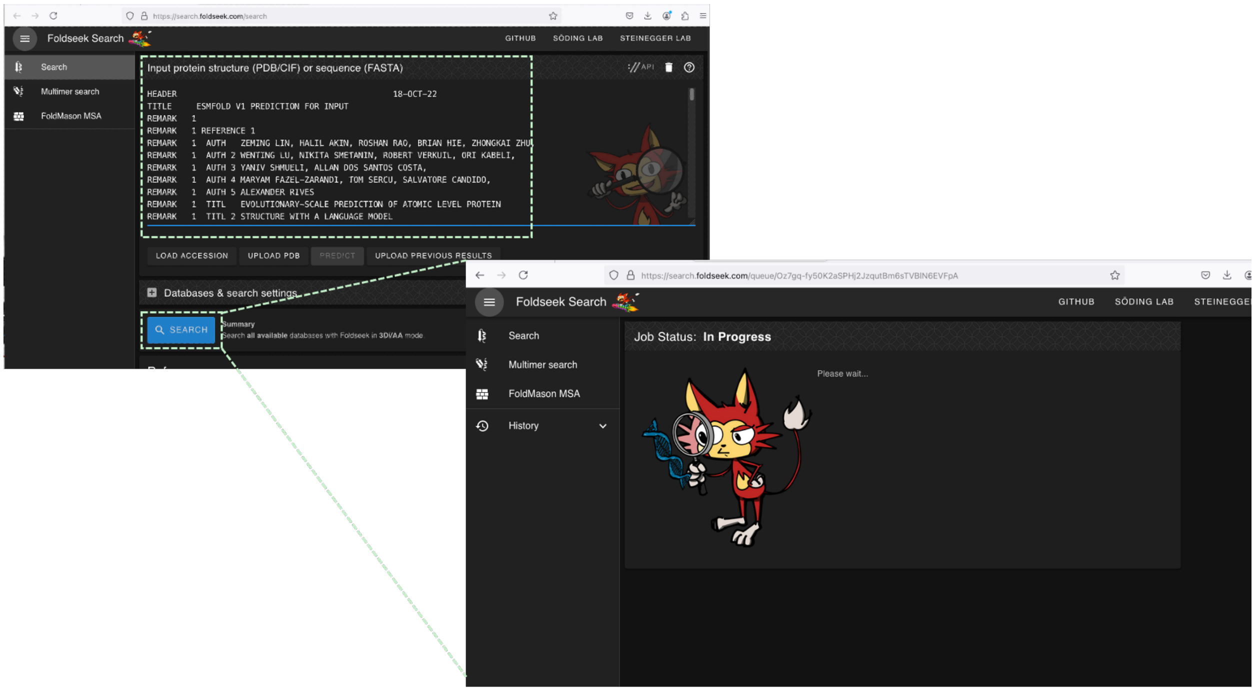


https://search.foldseek.com/search
Foldseek Search
GITHUB
SÖDING LAB
STEINEGGER LAB
Input protein structure (PDB/CIF) or sequence (FASTA)

```
MLPQVILSHLGSIGEVASTWLSENGYLSTQCPPLPKGNINYDIYMGYPEMFAWDKKRCVAYVSNLYNATV
STWDPYKGVVLDTIEFPGLSHAGDSASPNPLHASGIILRPDAYNPHTLEVVIDNGDAFYSDGFNVSGPDH
LMTIDLKTKEVTNKLRLNNGLYAGYADASLGPDGNTYVLGTYASNILRVTPEKEISTFYVADKLEPPRLY
GFTGIAHVGNAMIVPDNIIGQLIRFDVRDKVGTPVVIKQTPYHEFKTANVLHFPERYNDTILLVAENMTP
DHPFGGVSVYQDKTQQFNEVEFLGFLPSRLQNALTTSARQMTDRIYVVALPTDGANITVAGKSSRFEFQD
ITEELDLMIHPKSEGEVRDEI
```

“Doing the prediction”

Once the prediction completes, we can then search:



Input protein structure (PDB/CIF) or sequence (FASTA)

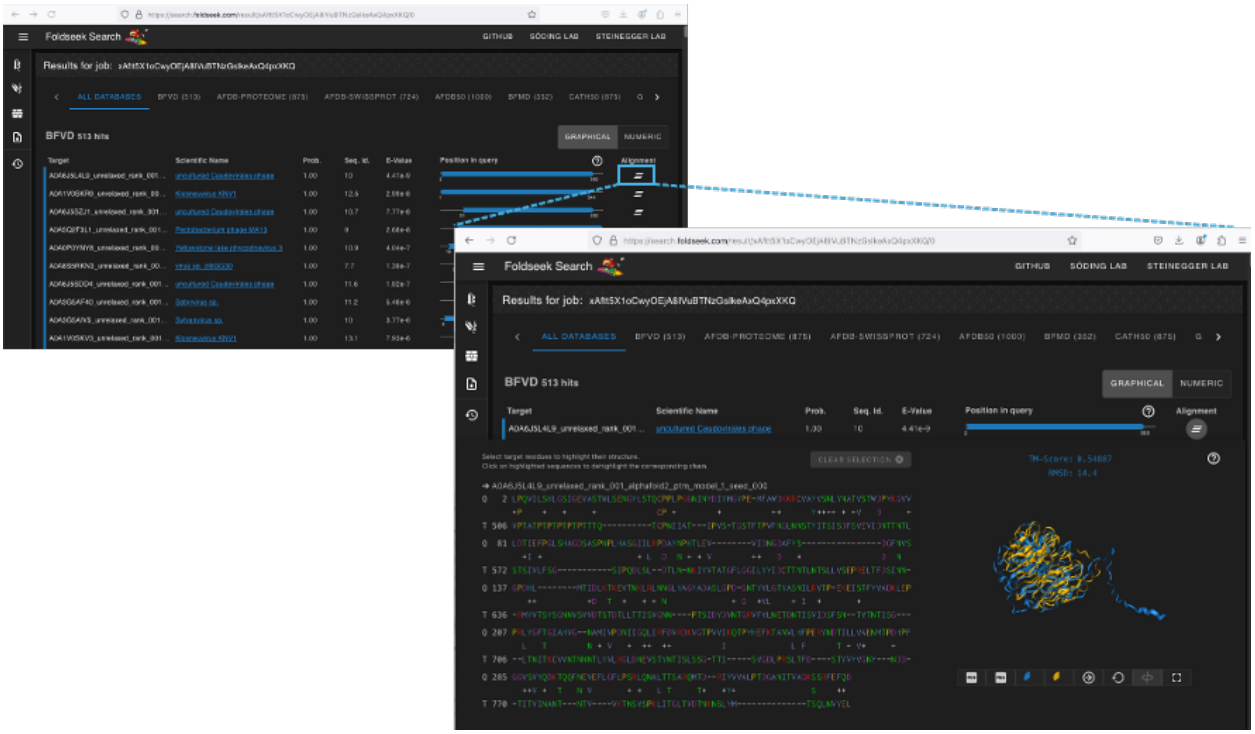
```
HEADER                                            18-OCT-22
TITLE     ESMFOLD V1 PREDICTION FOR INPUT
REMARK   1
REMARK   1 REFERENCE 1
REMARK   1  AUTH   ZEMING LIN, HALIL AKIN, ROSHAN RAO, BRIAN HIE, ZHONGKAI ZHU
REMARK   1  AUTH 2 WENTING LU, NIKITA SMETANIN, ROBERT VERKUIL, ORI KABELI,
REMARK   1  AUTH 3 YANIV SHMUELI, ALLAN DOS SANTOS COSTA,
REMARK   1  AUTH 4 MARYAM FAZEL-ZARANDI, TOM SERCU, SALVATORE CANDIDO,
REMARK   1  AUTH 5 ALEXANDER RIVES
REMARK   1  TITL   EVOLUTIONARY-SCALE PREDICTION OF ATOMIC LEVEL PROTEIN
REMARK   1  TITL 2 STRUCTURE WITH A LANGUAGE MODEL
```

Summary
Search all available databases with Foldseek in 3Di/AA mode

Job Status: In Progress

Please wait...

The hits from each database are made available, can now be explored, and saved.



Results for job: xAftt5X1oCwyOEjA8lVuBTNzGslkeAxQ4pxXKQ

BFVD 513 hits

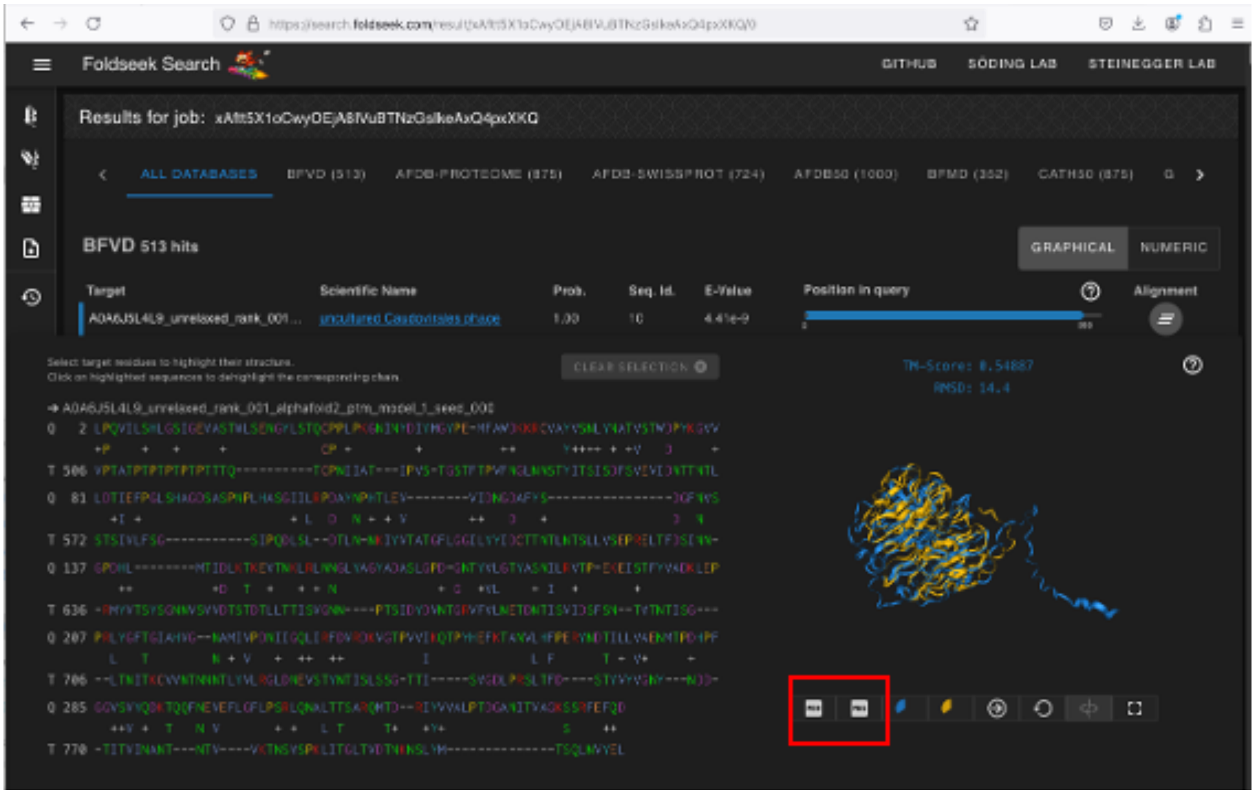
Target	Scientific Name	Prob.	Seq. Id.	E-Value
A0A6J5L4L9_unrelaxed_rank_001...	uncultured Caudovirales phage	1.00	10	4.41e-9
A0A1V0SKR9_unrelaxed_rank_00...	Klosneuvirus KNV1	1.00	12.5	2.08e-8
A0A6J5SZJ1_unrelaxed_rank_001...	uncultured Caudovirales phage	1.00	10.7	7.77e-8
A0A5Q2F3L1_unrelaxed_rank_001...	Pectobacterium phage MA13	1.00	9	2.08e-6
A0A0P0YMY8_unrelaxed_rank_00...	Yellowstone lake phycodnavirus 3	1.00	10.9	4.04e-7
A0A8S5RKN3_unrelaxed_rank_00...	virus sp. ct6S330	1.00	7.7	1.38e-7
A0A6J5SDD4_unrelaxed_rank_001...	uncultured Caudovirales phage	1.00	11.6	1.52e-7
A0A3G6AF40_unrelaxed_rank_001...	Sabrvirus sp.	1.00	11.2	5.46e-6
A0A3G6AIN5_unrelaxed_rank_001...	Sylvanvirus sp.	1.00	10	3.77e-6
A0A1V0SKV3_unrelaxed_rank_001...	Klosneuvirus KNV1	1.00	13.1	7.93e-6

Select target residues to highlight their structure.
Click on highlighted sequences to dehighlight the corresponding chain.

→ A0A6J5L4L9_unrelaxed_rank_001_alphafold2_ptm_model_1_seed_000

TM-Score: 0.54887
RMSD: 14.4

RESULTS CAN BE SAVED IN EITHER PDB OR IMAGE FORMAT (PNG) BY CLICKING THE BUTTONS BELOW THE SUPERPOSITION.



https://search.foldseek.com/result/xAftt5X1oCwyOEjA8lVuBTNzGslkeAxQ4pxXKQ/0

Foldseek Search

GITHUB SÖDING LAB STEINEGGER LAB

Results for job: xAftt5X1oCwyOEjA8lVuBTNzGslkeAxQ4pxXKQ

< ALL DATABASES BFVD (513) AFDB-PROTEOME (875) AFDB-SWISSPROT (724) AFDB50 (1000) BFMD (352) CATH50 (875) G >

BFVD 513 hits

GRAPHICAL NUMERIC

Target	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query	Alignment
A0A6J5L4L9_unrelaxed_rank_001…	uncultured Caudovirales phage	1.00	10	4.41e-9		

Select target residues to highlight their structure.
Click on highlighted sequences to dehighlight the corresponding chain.

CLEAR SELECTION

TM-Score: 0.54887
RMSD: 14.4

```
→ A0A6J5L4L9_unrelaxed_rank_001_alphafold2_ptm_model_1_seed_000
Q   2 LPQVILSHLGSIGEVASTWLSENGYLSTQCPPLPKGNINYDIYMGYPE-MFAWDKKRCVAYVSMLYNATVSTWDPYKGVV
      +P    +    +      +               CP +          +          ++     Y++++ + +V  D    +
T 506 VPTATPTPTPTPTPTTTQ----------TCPNIIAT---IPVS-TGSTFTPWFNGLMNSTYITSISDFSVEVIDNTTNTL

Q  81 LDTIEFPGLSHAGDSASPNPLHASGIILRPDAYNPHTLEV--------VIDNGDAFYS---------------DGFNVS
        +I +                   + L  D  N + + V        ++  D   +                  D N
T 572 STSIVLFSG-----------SIPQDLSL--DTLN-NKIYVTATGFLGGILYYIDCTTNTLNTSLLVSEPRELTFDSINN-

Q 137 GPDHL--------MTIDLKTKEVTNKLRLNNGLYAGYADASLGPD-GNTYVLGTYASNILRVTP-EKEISTFYVADKLEP
        ++           +D  T +   + + N                 + G  +VL    + I  +         +
T 636 -RMYVTSYSGNNVSVVDTSTDTLLTTISVGNN----PTSIDYDVNTGRVFVLNETDNTISVIDSFSN--TYTNTISG---

Q 207 PRLYGFTGIAHVG--NAMIVPDNIIGQLIRFDVRDKVGTPVVIKQTPYHEFKTANVL+FPERYNDTILLVAENMTPD+PF
        L   T         N + V   +  ++  ++             I              L F       T + V+     +
T 706 --LTNITKCVVNTNNNTLYVLRGLDNEVSTYNTISLSSG-TTI-----SVGDLPRSLTFD----STVVYVGNY---NDD-

Q 285 GGVSVYQDKTQQFNEVEFLGFLPSRLQNALTTSARQHTD--RIYVVALPTDGANITVADKSSRFEFQD
        ++V +  T   N V        + +   L T    T+   +Y+                S    ++
T 770 -TITVINANT---NTV----VKTNSVSPKLITGLTVDTNKNSLYM-------------TSQLMVYEL
```