

Use of the **HHPRED** &  
**AlphaFold** servers to predict  
a protein 3D structure;  
comparison of obtained  
models with **PyMol**  
& model quality check with  
**QMEAN** and **ProQ2**

# HHPRED

Go to

<https://toolkit.tuebingen.mpg.de/tools/hhpred/>

HHpred is a method based on the pairwise comparison of *profile* hidden Markov models (HMM-HMM alignment) for remote protein homology detection and 3D structure prediction.

(Söding et al 2005 NAR **33**:W244)

# HHPRED *input*

Go to:

<https://toolkit.tuebingen.mpg.de/tools/hhpred/>

- Copy the protein FASTA sequence of horse **AQP11** (Uniprot ID: **F6S3G9**) in the input box
- Select the database **PDB\_mmCIF70\_lastdate** (default DB)
- Click on **Submit**
- Then click either on **Load existing job** or on **Start job anyway** (in this case you'll have to wait)

Go to

<https://toolkit.tuebingen.mpg.de/tools/hhpred/>

HHPRED

*Input-1*

The screenshot displays the HHPRED web interface. At the top, there is a navigation menu with options: Search, Alignment, Sequence Analysis, 2ary Structure, 3ary Structure, Classification, and Utils. Below this, a sub-menu includes HHblits, HHpred, HMMER, PatternSearch, and ProtBLAST/PSI-BLAST. The main header shows the MPI Bioinformatics Toolkit logo and a search bar. A table on the left lists recent jobs with columns for ID, Date, and Tool. The main content area is titled 'HHpred' and shows a job ID of 8830209, created 14 minutes ago. The 'Input' tab is active, displaying a text area for entering a protein sequence or alignment. Below the text area, there are options to 'Align two sequences/MSAs' (disabled), 'Select structural/domain databases' (set to 'PDB\_mmCIF70\_31\_Jul'), and 'Select proteomes' (set to 'Select options'). A 'Resubmit' button is visible at the bottom right.

ID	Date	Tool
8830209		HHPRED
9698875		HHPRED

Job ID: 8830209, Created: 14 minutes ago

Input

Enter a protein sequence/multiple sequence alignment in A3M/CLUSTAL/FASTA/STOCKHOLM format. For template-based homology modeling using MODELLER, select the PDB\_mmCIF70 or PDB\_mmCIF30 database.

Align two sequences/MSAs

Select structural/domain databases: PDB\_mmCIF70\_31\_Jul

Select proteomes: Select options

8830209\_2 ✓ Resubmit

# HHPRED

## *Output-1*

Go to

<https://toolkit.tuebingen.mpg.de/tools/hhpred/>

- Have a look at the obtained results
- How many hits (3D structures of homologous proteins) are proposed?
- How many hits with a very good match were found?
- Select all the hits with an E-value better (lower) than  $1 \times 10^{-10}$
- Visualize the alignment with the last selected hit (template for modeling)
- Click on **Model using selection**

# HHpred *input-2*

The screenshot displays the HHpred web interface. At the top, there are navigation tabs for Search, Alignment, Sequence Analysis, 2ary Structure, 3ary Structure, Classification, and Utils. Below these are sub-tabs for HHblits, HHpred, HMMER, PatternSearch, and ProtBLAST/PSI-BLAST. The main header shows the MPI Bioinformatics Toolkit logo, a search bar, and a list of recent jobs with columns for ID, Date, and Tool. The current job is highlighted in green.

The HHpred results page shows the following summary:

- Number of Hits: **71**
- Query MSA diversity (Neff): **10.4752**
- Detected sequence features: **■ Transmembrane segment(s)**

The 'Model using selection' button in the top navigation bar is highlighted with a blue box and a black arrow pointing to it. Below the summary is a 'Visualization' section with a slider ranging from 40 to 265. A 'Resubmit Section' button is located at the top right of the visualization area. The visualization shows 12 red horizontal bars representing transmembrane segments, with their corresponding PDB IDs listed on the right:

- 3CN5\_R
- 3C02\_R
- 6KXM\_C
- 3IY2\_R
- 209G\_R
- 1L0F\_R
- 1J4N\_R
- 40J2\_X
- 2229\_R
- 5I32\_R
- 3Z0J\_R
- 2860\_R

# HHPRED

## *Input-2*

Go to

<https://toolkit.tuebingen.mpg.de/tools/hhpred/>

- Take note of the steps performed by the software
- Have a look at the obtained multiple sequence alignment (MSA) between our query (P<sub>1</sub>;UKNP) and the selected hits
- Now click on **Forward to MODELLER**
- You will now be asked to enter a **MODELLER-key**

HHPRED

*Output-2*

Go to

<https://toolkit.tuebingen.mpg.de/tools/hhpred/>

- Take note of the steps performed by the software for modeling
- Click on **3D Structure** to visualize the 3D model
- Now click on **Download PDB File**

# HHPRED *Output-2*

MPI Bioinformatics Toolkit

Search Alignment Sequence Analysis 2ary Structure 3ary Structure Classification Utils

MODELLER SamCC

Job ID: 5283124, Parent Job ID: 2003711, Created: 1 minute ago

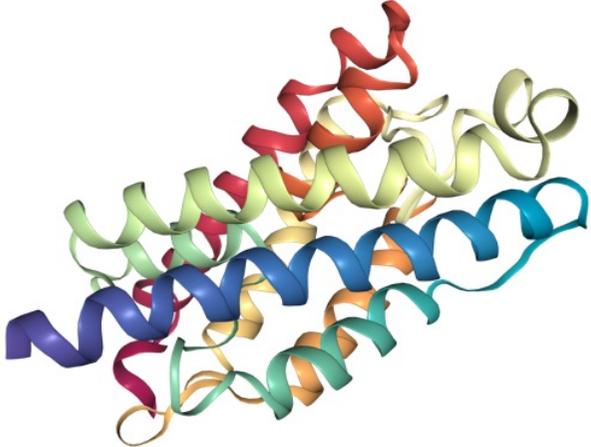
**MODELLER** ⓘ

ID	Date	Tool	
5283124		MODL	×
2003711		HTMP	×
3180450		MODL	×
4614495		HTMP	×
9305044		HHPR	×

Input 3D Structure

[Download PDB File](#)

Download the obtained model



# AlphaFold

It is an advanced deep learning algorithm, based on neural networks.

It is able to predict protein structures with atomic accuracy even in cases in which no similar structure is known.

(Jumper et al. 2021 Nature **596**:583)

# AlphaFold DB

## Input

contains 214,683,829 structures, including 48 complete proteomes (at nov 14<sup>th</sup> 2022)

Go to

<https://alphafold.ebi.ac.uk>

AlphaFold Protein Structure Database

Home About FAQs Downloads

# AlphaFold Protein Structure Database

Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism BETA Search

Examples: Free fatty acid receptor 2 At1g58602 Q5V5... E. coli Help: AlphaFold DB search help

Feedback on structure: Contact DeepMind

Insert the Uniprot code F6S3Gg of aqp11 of Equus caballus

- Click on **Search**

# AlphaFold DB

*output*

Go to

<https://alphafold.ebi.ac.uk>

- Select the entry corresponding to your protein
- Click on **Download PDB File**
- Have a look at the Information relative to “Experimental structures”
- In the **3D viewer**, have a look at the “**Model Confidence**”
- Take a look at the protein regions (residues) with the lowest level of confidence
- Is there any part of the model predicted with a “very low” confidence?

# AlphaFold DB *output*

Go to

<https://alphafold.ebi.ac.uk>

EMBL-EBI Services Research Training About us EMBL-EBI

AlphaFold Protein Structure Database Home About FAQs Downloads

Search for protein, gene, UniProt accession or organism BETA Search

Examples: [Free fatty acid receptor 2](#) [At1g58602](#) [Q5VSL9](#) [E. coli](#) Help: [AlphaFold DB search help](#)

## Aquaporin-11

AlphaFold structure prediction

Download [PDB file](#) [mmCIF file](#) [Predicted aligned error](#)

**NEW** Feedback on structure [Looks great](#) [Could be improved](#)

### Information

Protein	Aquaporin-11
---------	--------------

Download the file in format of your choice  
(we'll use the PDB format)

# Models visualization with PyMol

Go to

<https://pymol.org/edu/>

- Register yourself as a student for Educational-Use-Only PyMOL Builds: insert your data then click on **Continue**
- Once obtained the license as a student, **download** PyMOL Builds

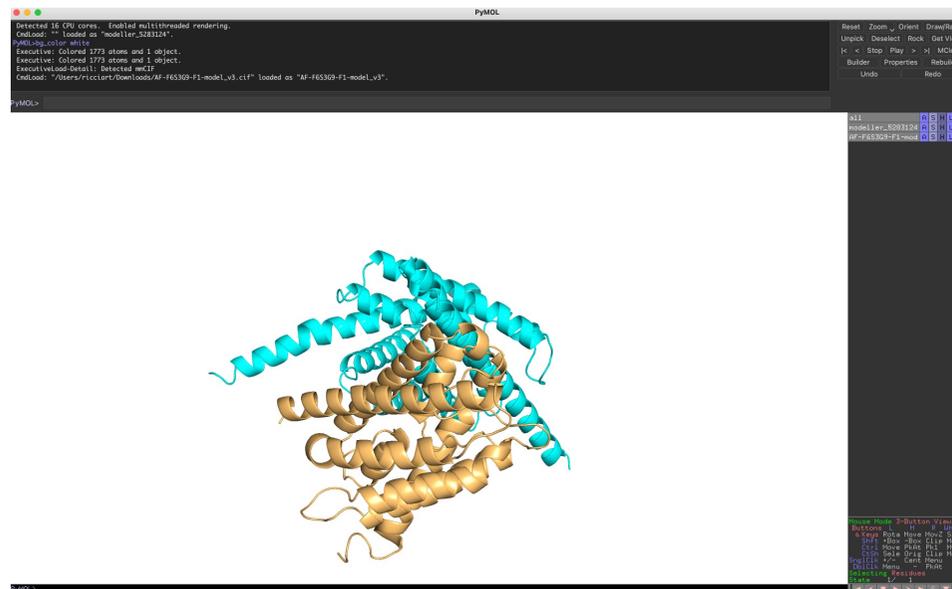
*Username:*

*Password:*

- PyMOL is a software to visualize the 3D-structure of molecules and to perform some minor manipulations on them (single amino acid mutation, etc.)

# Models visualization with PyMol

- Load in PyMol the 3D models obtained in previous steps, by sleceting from the menu: **File – open – select file**
- Superimpose them with the command line:  
**align namemodel<sub>1</sub>, namemodel<sub>2</sub>**
- Color the AlphaFold model by B-factor (i.e. the confidence score)
- Check if regions with a low confidence score diverge from corresponding regions predicted by HHPRED



# Quality check with QMEAN

*Input*

Go to

<https://swissmodel.expasy.org/qmean/>

- Upload the PDB file for the model(s) by clicking on **Select Coordinate File**
- Select **QMEANDisCo** as the method
- Now click on **Submit**

Quality check  
with **QMEAN**  
*Output*

Go to

<https://swissmodel.expasy.org/qmean/>

- Wait for the output. Visualize it. How is on average the confidence score for the model?
- What are the low-confidence regions of the model?

# Quality check with ProQ2

*Input*

Go to

<http://bioinfo.ifm.liu.se/proq2/index.php>

- Upload the PDB file for the model(s) by clicking on **Choose File**
- Keep the default options (“Return model with quality in B-factor column”)
- Digit your email address in the **Email** box
- Now click on **Submit**

# Quality check with ProQ2

## *Output*

Go to

<http://bioinfo.ifm.liu.se/proq2/index.php>

- The output can take some time, you will be notified by email
- When ready, have a look at it. Is the confidence score generally high for the model?
- Compare with the QMEAN output