

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**

HHRED

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 bar with links for Search, Alignment, Sequence Analysis, 2ary Structure, 3ary Structure, Classification, and Utils. Below this, a sub-navigation bar lists HHblits, HHpred, HMMER, PatternSearch, and ProtBLAST/PSI-BLAST. The main header includes the MPI Bioinformatics Toolkit logo and a search bar. A table on the left lists job IDs and tools, with two entries for HHPR. 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 multiple sequence alignment in A3M/CLUSTAL/FASTA/STOCKHOLM format. Below the text area, there is a toggle for 'Align two sequences/MSAs' and a section for 'Select structural/domain databases' with a dropdown menu showing 'PDB_mmCIF70_31_Jul'. To the right, there is a 'Select proteomes' dropdown menu. At the bottom right, there is a 'Resubmit' button and a status indicator for job 8830209_2.

MPI Bioinformatics Toolkit

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

HHblits HHpred HMMER PatternSearch ProtBLAST/PSI-BLAST

HHpred ? Job ID: 8830209, Created: 14 minutes ago

Input Parameters Results Raw Output Probability Plot Query Template MSA Query MSA

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 1×10^{-10}
- Visualize the alignment with the last selected hit (template for modeling)
- Click on **Model using selection**

HHpred

input-2

MPI Bioinformatics Toolkit

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

HHblits HHpred HMMER PatternSearch ProtBLAST/PSI-BLAST

HHpred ? Job ID: 9305044, Created: 1 day ago

Input Parameters Results Raw Output Probability Plot Query Template MSA Query MSA

Vis Hits Aln Select All Forward Forward Query A3M Model using selection Download HHR Color Seqs Wrap S

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

Visualization

40 265 Resubmit Section

3CN5_R
3CB2_R
6KXM_C
3IYZ_R
209G_R
1L0F_R
1J4N_R
40J2_X
2229_R
5I32_R
3Z0J_R
2860_R

HHRED

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 (P₁;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

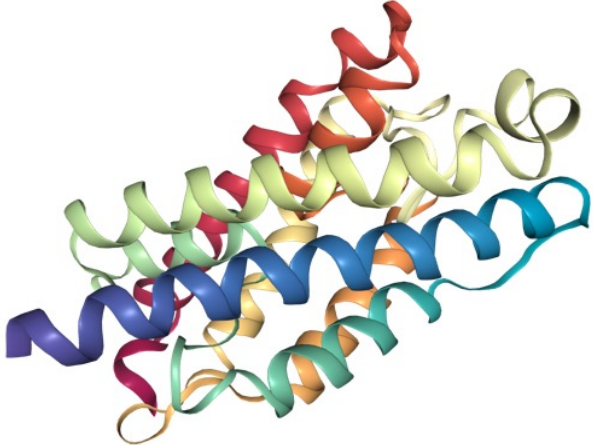
MODELLER ?

| 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 14th 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 Q5V5S1 E. coli Help: AlphaFold DB search help

Feedback on structure: Contact DeepMind

Insert the Uniprot code F6S3G9 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>

The screenshot shows the AlphaFold Protein Structure Database interface. At the top, there is a navigation bar with links to Home, About, FAQs, and Downloads. Below this is a search bar with the placeholder text "Search for protein, gene, UniProt accession or organism" and a "Search" button. A "BETA" label is also present. Below the search bar, there are examples of search terms: "Free fatty acid receptor 2", "At1g58602", "Q5VSL9", and "E. coli". A "Help" link points to "AlphaFold DB search help".

The main content area displays the search results for "Aquaporin-11". The title "Aquaporin-11" is in large, bold letters, followed by "AlphaFold structure prediction". Below this, there is a "Download" section with three buttons: "PDB file", "mmCIF file", and "Predicted aligned error". A callout box with a black border and white text points to these buttons, stating: "Download the file in format of your choice (we'll use the PDB format)".

Below the download buttons, there is a "Note" stating: "We have recently updated the PAE JSON format, please refer to our [FAQ](#) for a description of the updated format." Below the note, there is a "Feedback on structure" section with a "NEW" label and two buttons: "Looks great" and "Could be improved".

At the bottom, there is an "Information" section with a table showing the protein name "Aquaporin-11".

| Information | |
|-------------|--------------|
| Protein | Aquaporin-11 |

Models visualization with PyMol

Go to

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

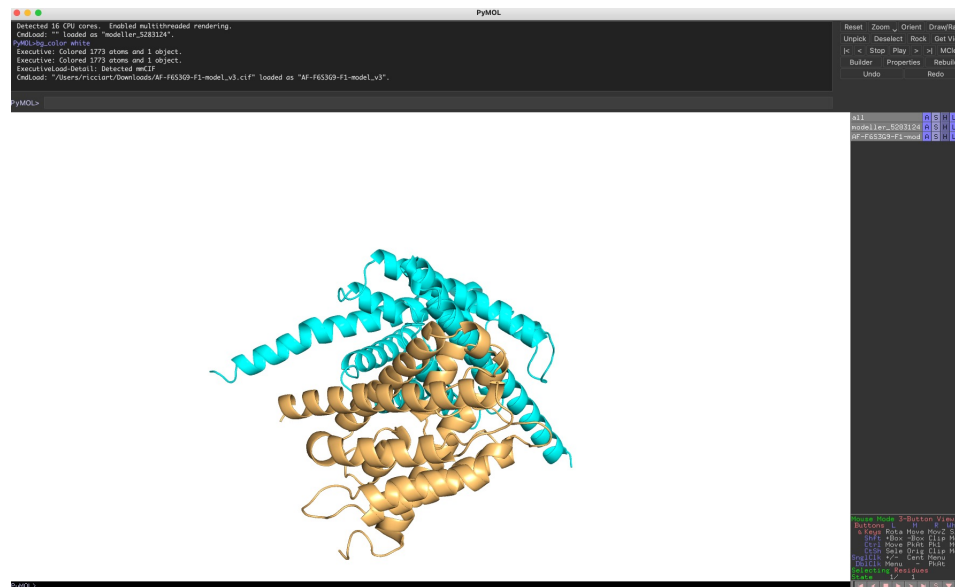
- Register yourself as a student for Educational-Use-Only PyMOL Builds: insert your data then click on **Continue**
- Once obtained the license as a student, **download** PyMOL Builds

Username: XXXXXXXXXXXX
Password: XXXXXXXXXXXX

- 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 namemodel1, namemodel2
- 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