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

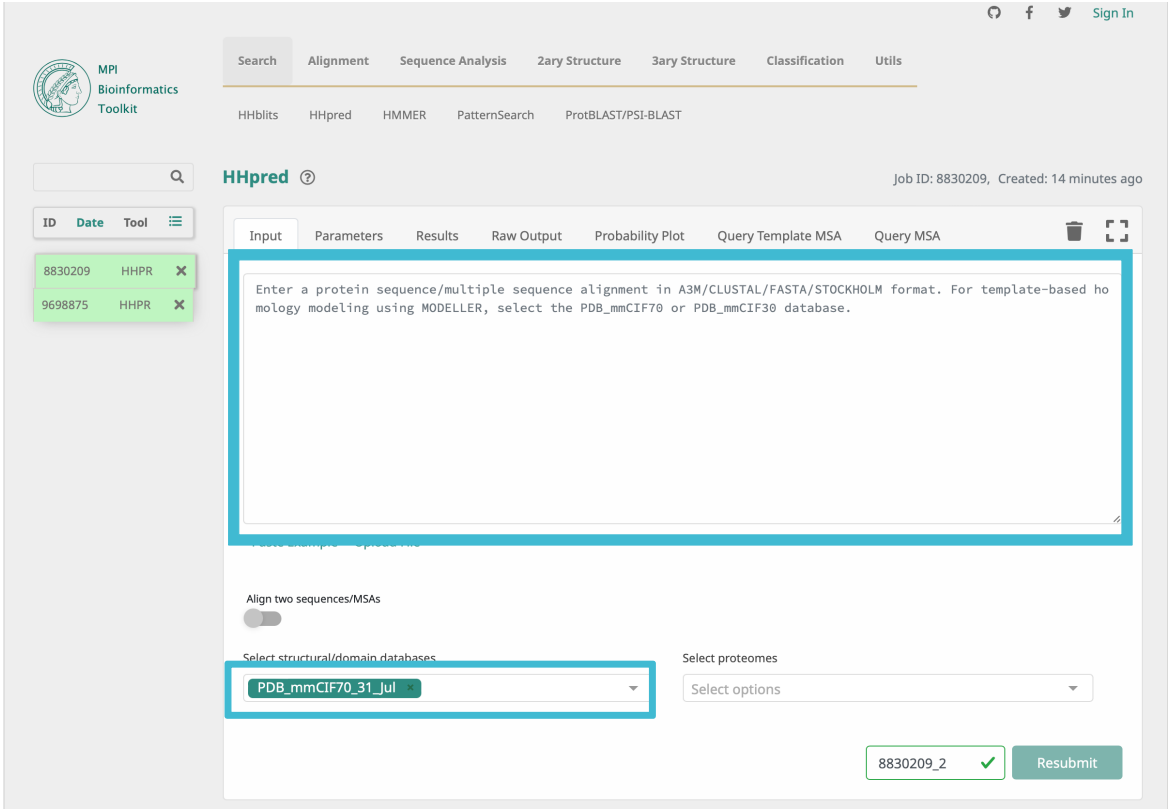
Go to:
https://toolkit.tuebingen.mpg.de/tools/hhpred/

- 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)

HHPRED

Input-1

Go to
https://toolkit.tuebingen.mpg.de/tools/hhpred/



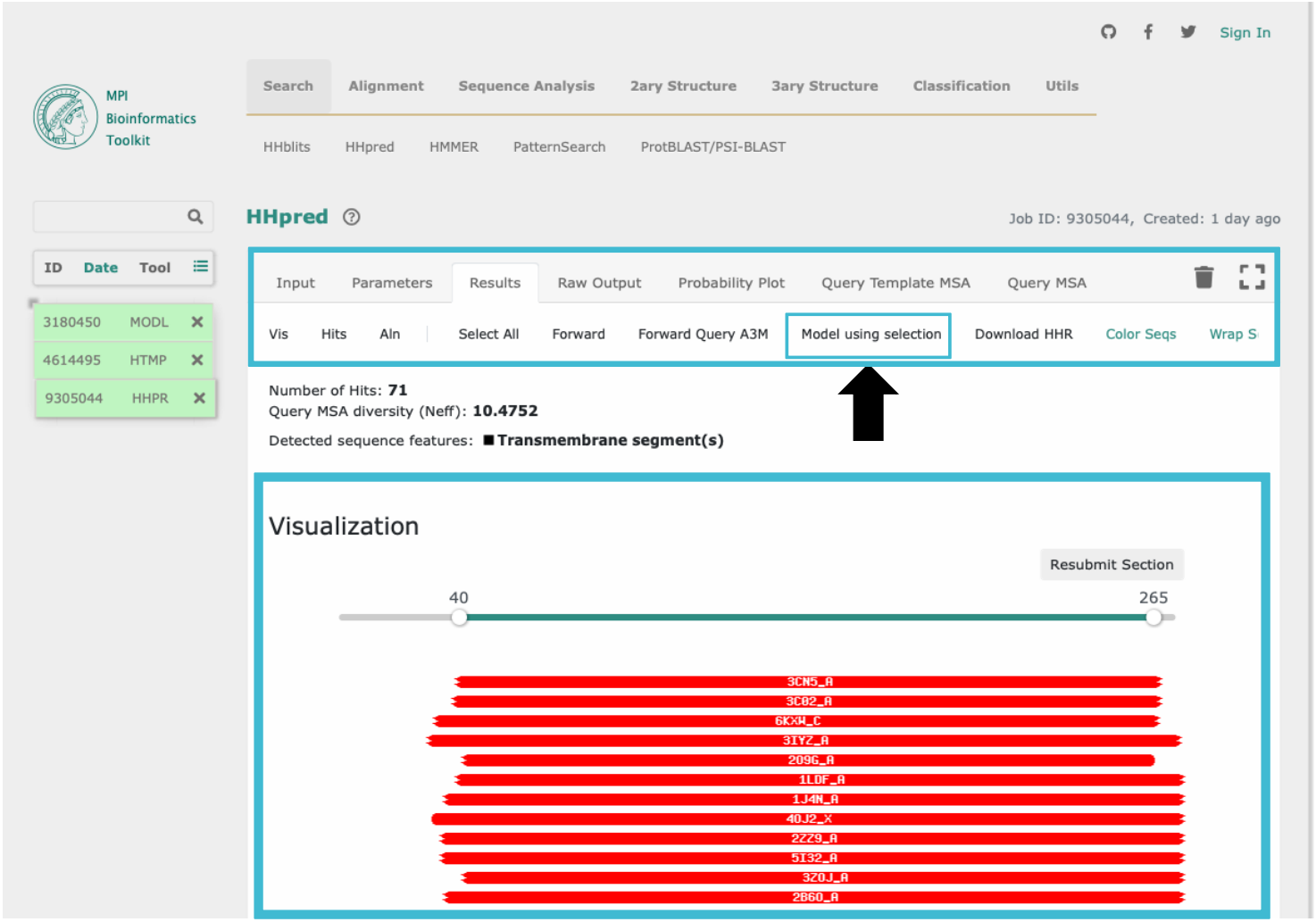
HHPRED

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 $1x10^{-10}$
- Visualize the alignment with the last selected hit (template for modeling)
- Click on **Model using selection**

HHPRED
input-2



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 (P1;UKNP) and the selected hits
- Now click on **Forward to MODELLER**
- You will now be asked to enter a **MODELLER-key**

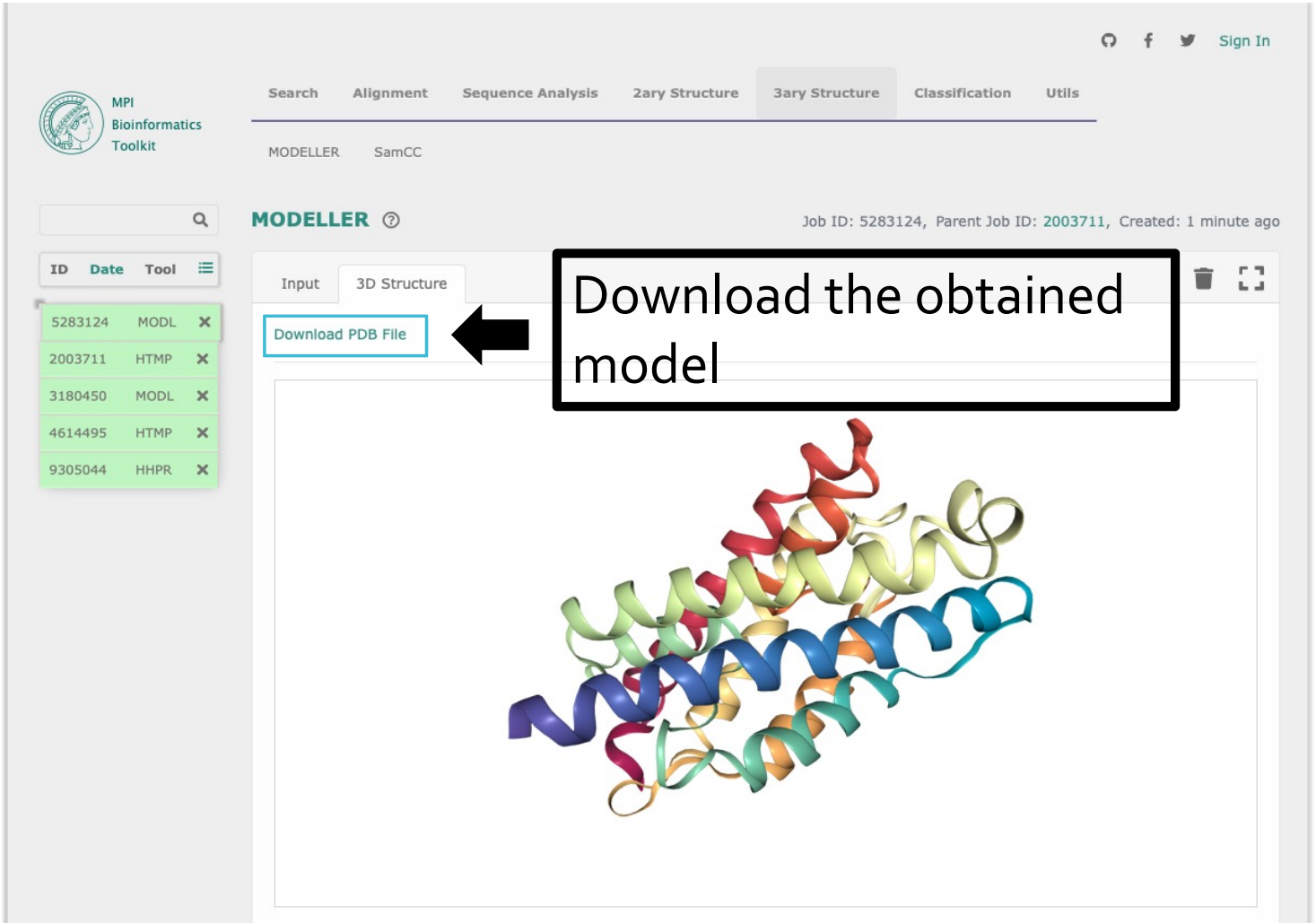
HHPRED

Ouput-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
Ouput-2



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.

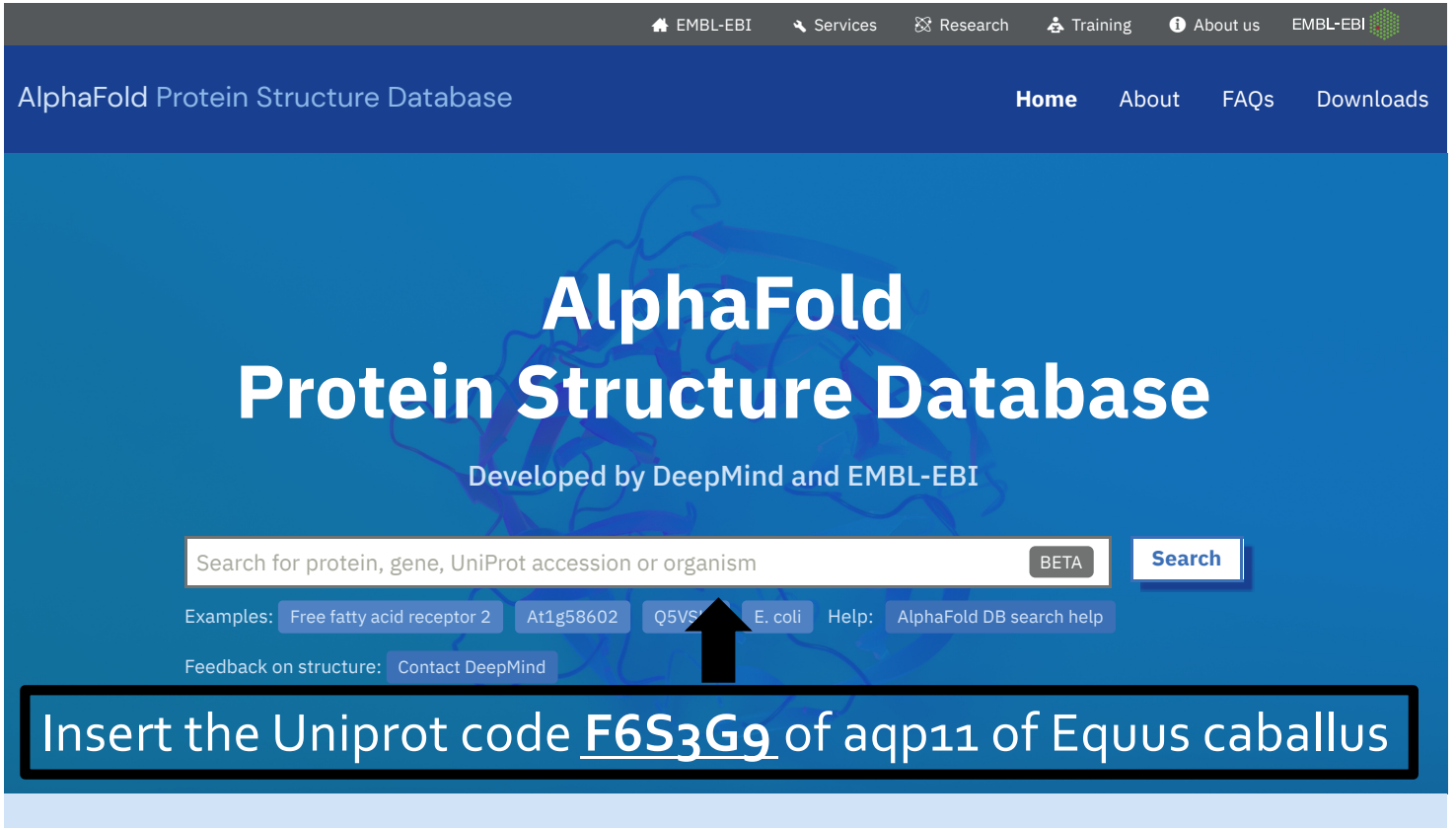
(Jumper et al. 2021 Nature **596**:583)

AlphaFold DB

Input

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

Go to
https://alphafold.ebi.ac.uk



- Click on **Search**

AlphaFold DB

output

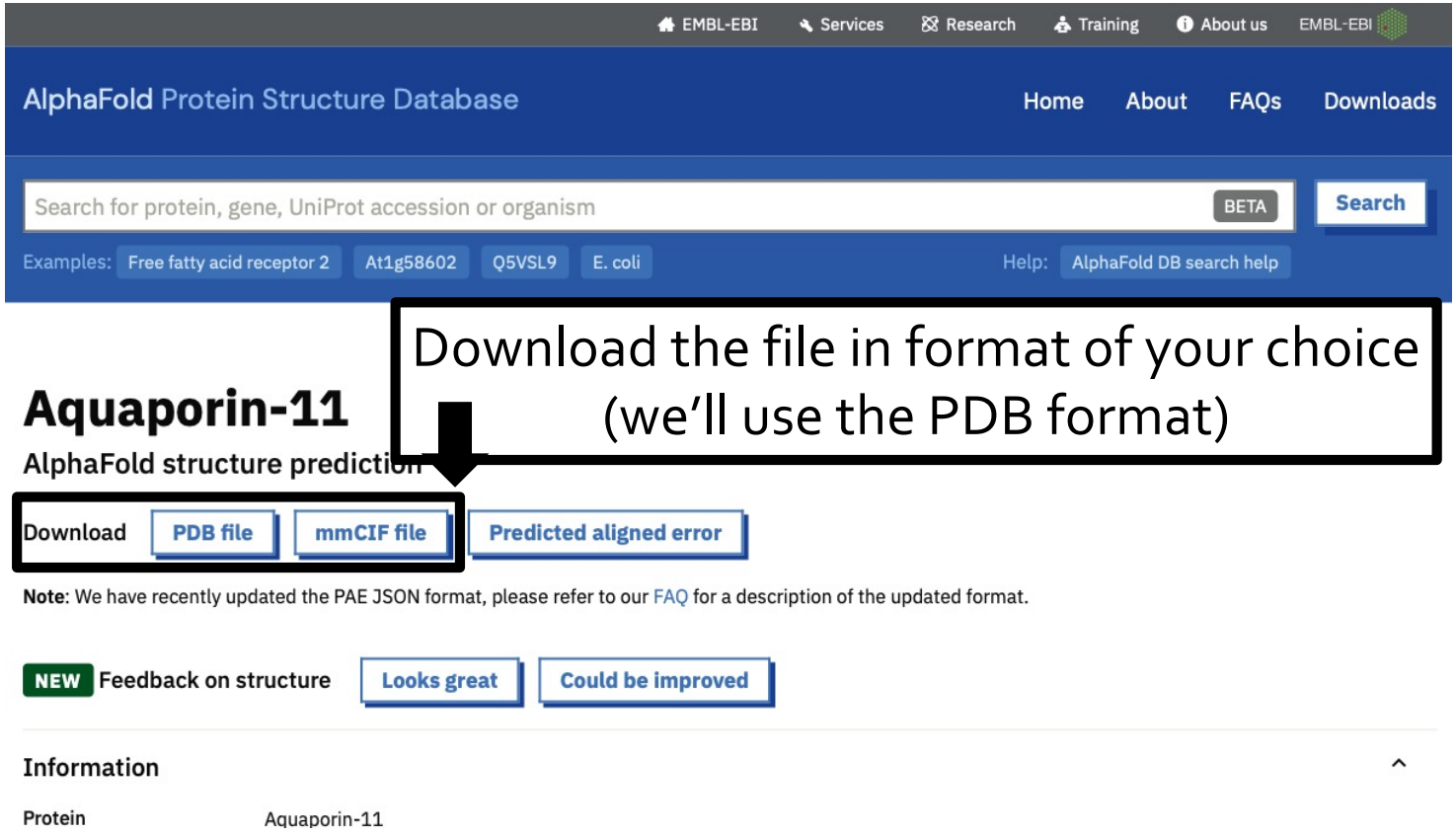
Go to
https://alphafold.ebi.ac.uk

- 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



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

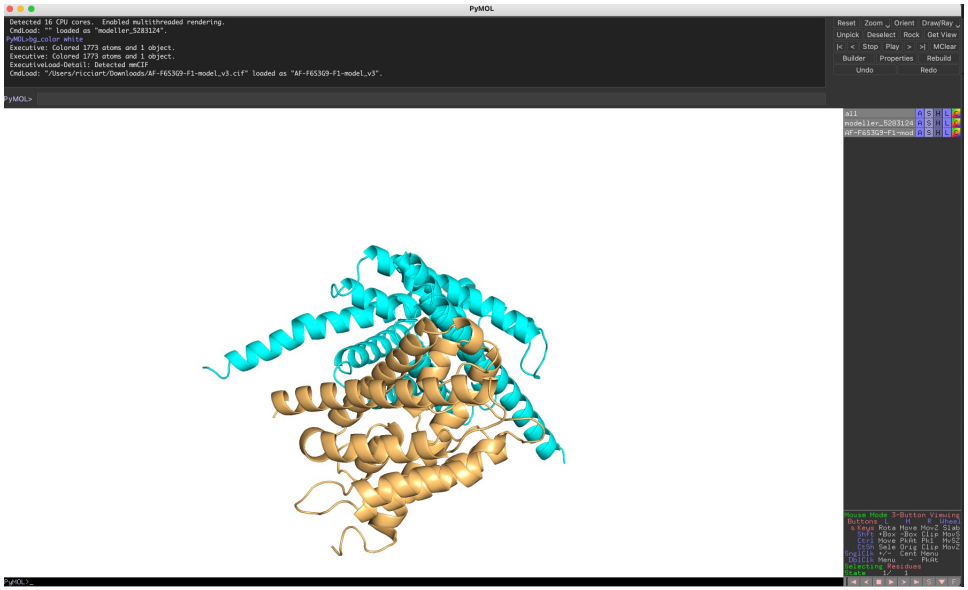
  Username: Xxxxxxxxxx
  Password: Xxxxxxxxxx

- 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