# Some steps I did as preparation

# Creating the model

### Question 1

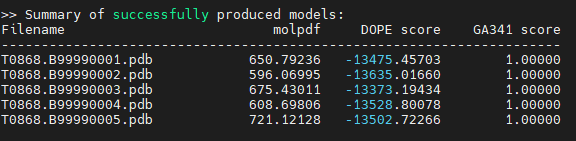
1. Find sequence of protein T0868 on CASP site (<https://predictioncenter.org/casp12/target.cgi?id=12&view=all>)
2. Go to HHPred server en copy paste the sequence from step 1. Use the default HHpred parameters but changing the ”MSA generation method” command into PSI-BLAST=>nr70. (<https://toolkit.tuebingen.mpg.de/tools/hhpred>)
3. HHpred found a already excisting protein in there database. I loaded that protein and got here: <https://toolkit.tuebingen.mpg.de/jobs/6527440>
4. I selected the first hit and selected ‘Model using selection’:



1. Got here and copied the pir formatted text (<https://toolkit.tuebingen.mpg.de/jobs/5159171>)
2. On PDB I downloaded the 5j4a.dpb fiel
3. Changed the following things in the alignment.ali file:
   1. Fields 3-6 of the sequence, should say something about the start and end of the residue and chain. But for the sequence we got from CASP, this is unknown. So I removed it
   2. The name of the 5J4A is changed to 5j4a.pdb. Also, the starting and ending of the residues and chains (201 :C:316 :C ) is checked in the 5j4a.pdb file, and this is correct, so no alterations were made in fields 3-6 of the structure.

Then, to model

1. I started MobaXterm and created the folder HHpredModel and uploaded the 5j4a.pdb and alignment.ali files.
2. Created the file buid\_model.py in the same folder. Edited the variables: alnfile, knowns and sequence.



The dope Score will tell us about the accuracy and efficiency of the Model. The model with the minimum score is the best model (don’t know why). In our case this is model T0868.B99990002.pdb with DOPE score of -13635.

### Question 2

Chimera: <https://www.cgl.ucsf.edu/chimera/>



Gold is our model created with MODELLER. Blue is the template structure. So no big gaps.

So the new file alignment.ali is made with the following:

C; bladiebla

>P1;T0868

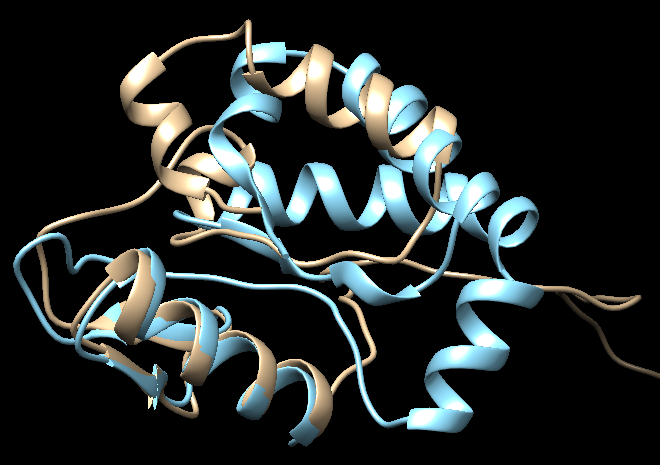
sequence:T0868: . : . : . : : : : :

MGASSGSNISASNGSSSPTTIVASNPVDLNAFDRLNVVDPAVGKFRPGEAGAAAELENYLGGTLQRAPQGSSVDFVFSSGPNN---------------------------------DHAAAADFVPLASRFLSEANKTLLVKAIGNLPQKLQAKIILIK\*

>P1;5j4a

structure:5j4a.pdb:201 :C:316 :C::Burkholderia pseudomallei:5J4:

---------------------------------------------RPGEAGAAAELENYLGGTLQRAPQGSSVDFVFSSGPNNGKTVDFMLTPDTVAQAAKINQFFDKNLNNFMNTLSDHAAAADFVPLASRFLSEANKTLLVKAIGNLPQKLQAKIILIK\*



This image shows that the model just skips a whole part of the sequence, the gaps are deleted and it looks like the sequence before and after the gap are merged.

# Scoring you model

### Question 3

The “-aa1:begin:end” and “-aa2:begin:end” parameters should give the ranges that you know are corresponding between the reference structure and your model structure… ?? (don’t know how to do this.. )

I created the following files:

* + - shared-data-sets/LGA/LGA\_package\_src/MOL2/5j4a\_C.T0868\_C
    - shared-data-sets/LGA/LGA\_package\_src/TMP/lrn208

How do you run the /groups/Shared-Data-Sets/LGA/lga 5j4a\_C.T0868\_C -4 > TMP/lrn208.txt command in MobaXterm?

### Question 4

# Structural comparison

### Question 5

# Pairwise sequence alignment

### Question 6

### Question 7

# Paper by Forrest et al. (2006)

### Question 8

### Question 9

# Contributions

### Question 10