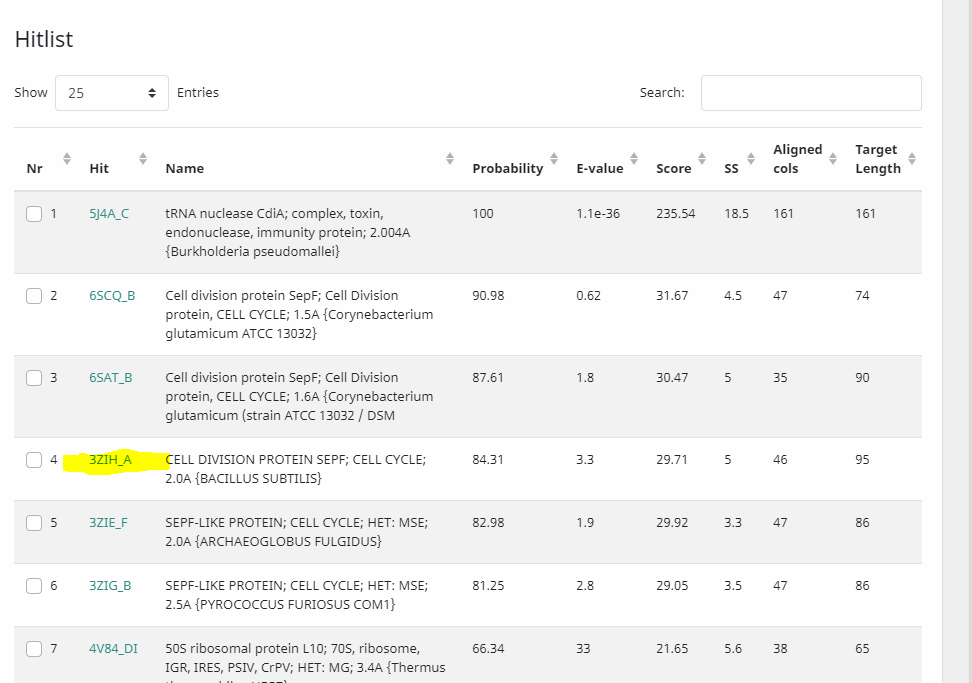
# Creating the model

### Question 1

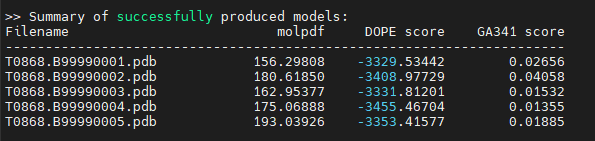
1. Find sequence of protein T0868 on CASP site (<https://predictioncenter.org/casp12/target.cgi?id=12&view=all>)
2. Go to HHPred server en copy paste the sequence from step 1. Use the default HHpred parameters but changing the ”MSA generation method” command into PSI-BLAST=>nr70. (<https://toolkit.tuebingen.mpg.de/tools/hhpred>)
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 yellow hit and selected ‘Model using selection’:



1. Got here and copied the pir formatted text (<https://toolkit.tuebingen.mpg.de/jobs/5159171>) into the alignment.ali file
2. 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.
3. On PDB I downloaded the 3zih.pdb file

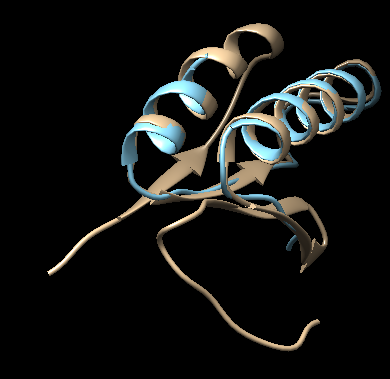
Then, to model

1. I started MobaXterm and created the folder HHpredModel and uploaded the 3zih.pdb and alignment.ali files.
2. Created the file build\_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.B99990004.pdb with DOPE score of -3455.46704 !

### Question 2



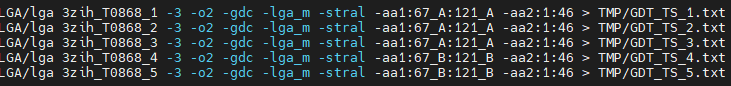
Long ga

# Scoring you model

### Question 3

I created the 3zih\_T0868\_1 … \_5 files in the MOL2 folder…

Did used this command line for 1 till 5:



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Range solution | Range model | DOPE | GDT\_TS |
| 1 | 67\_A:121\_A | 01:46 | -332.953.442 | 37.500 |
| 2 | 67\_A:121\_A | 01:46 | -340.897.729 | 38.587 |
| 3 | 67\_A:121\_A | 01:46 | -333.181.201 | 37.500 |
| 4 | 67\_B:121\_B | 01:46 | -345.546.704 | 38.587 |
| 5 | 67\_B:121\_B | 01:46 | -335.341.577 | 39.130 |

The greener, the more the model structure looked like the solution according to the measure. So this differs pretty much per model!

### Question 4

So when looking at the GDT\_TS\_4.txt file.. you can tell that from the model, atoms 22 till 33 have a low GDC\_all score and therefor are modelled good.. This part I just a alpha helix in the model, see in Chimera..

Morgen even vragen… waar moet ik naar kijken in de tabel en in het model? Lijkt niet echt overeenkomsten in te zitten..

# Structural comparison

### Question 5

# Pairwise sequence alignment

So first we made the alignment.ali file with HHpred. HHpred is not a sequence alignment, its is based on the pairwise comparison of profile hidden Markov models (HMMs). Whereas most conventional sequence search methods search sequence databases such as UniProt or the NR, HHpred searches alignment databases, like Pfam or SMART.

### Question 6

1. I used the LOCAL (I think local is needed because the alignment from the slide looks also local) Water EMBOSS pairwise sequence alignment (random choice) to generate the alignment: https://www.ebi.ac.uk/Tools/psa/emboss\_water/ . Input files were seq\_3zih and seq\_T0868. I choose the pearson/fasta output..
2. I pasted the result in the alignment.ali file.. this needs to be converted to pir format. Also the AA for the template sequence needs to be checked.

### Question 7

# Paper by Forrest et al. (2006)

### Question 8

### Question 9

# Contributions

### Question 10