**Minutes from meeting 2023.09.12**

* We looked at the alignment and tree of the seed sequences (the ones studied by Caio, and the ones in Caio’s structure table):
  + There are some almost identical protein structures in the list Caio sent. We have to check the original papers and choose the wildtype only.
  + The additional chains in 2Y9X and 2Y9W should be removed. They look like something else.
  + 4OUA is weird – it has two chains where one part is identical, but one has an extra C-terminal domain
* We need to decide whether we should blast the whole protein or only the pfam domain. And whether we should cluster based on the whole protein or only the pfam domain.
* We discussed the clustering methods. We can start with phylogenetic trees. If the proteins are too diverse, we can use Sequence Similarity Networks (SSNs). Maybe make a figure like Figure 2 in GH16 paper (Viborg, 2019).

**Ida tasks:** Write draft to project description (at least the bioinformatics part). Look into the structure papers and clean the seeds (remove chains). Make new alignment and tree. Get taxonomy information from seeds. Have a look at “phylogeny” papers. Maybe start making setup for blast.

**Jane tasks:** Check if she has access to Ida’s phdigital.