**Minutes**

11/02/23

* Discussed generating PSSMs.
  + Using HHblits, which is quicker and more up to date than PSI-BLAST.
  + Generating them via ssh to Kevin’s machine.
  + PSSMs as inputs should give more competitive results.
* Discussed RNN Notebook.
  + Showed RayTune (hyperparameter tuning).
  + Discussed validation data.
    - Cross validation.
    - As the validation dataset is smaller there can be a larger variability of sequences in this set, which is a reason the validation curve can show high loss spikes at some training iterations.
* Discussed length of protein sequences.
  + My smallest sequence is 19 amino acids long, my largest 34350.
  + This is ok, and something to consider in the evaluation.
  + ESM (Facebook’s protein language model) removed sequences that were above a given constraint for sequence length.
* Discussed dissertation.
  + I should include more research that has been done for protein disorder prediction using deep learning approaches alongside the other approaches not using deep learning.
* Discussed presentation of the Django server.
  + I will check that the server works and add finishing touches if necessary.

Goals for this week:

* Finish writing about current approaches in the form of a literature review.
* Write parts of the methodology. How I approached the problem.
* Generate and use PSSMs.