

Lab 11:

Qingyun Zheng (A16338254)

Table of contents

Background	1
AlphaFold	1
Interpreting/ Analyzing AF results in R	2

Background

We saw last day that the PDB has 209,886 entries (Oct/Nov 2025). UniprotKB (protein sequence database) has 199,579,901 entries.

```
209886/199579901 *100
```

```
[1] 0.1051639
```

So the PDB has only 0.1% coverage of the main protein sequence database.

Enter AlphaFold database (AFDB) <https://alphafold.ebi.ac.uk/> that attempts to provide computed models for all sequences in UniProt

“AlphaFold DB provides open access to over 200 million protein structure predictions to accelerate scientific research.”

AlphaFold

AlphaFold has three main outputs:

- The predicted coordinates (PDB files)
- A local quality score called **pIDDT** (one for each amino-acid)
- A second quality called **PAE** (predicted aligned error, for each pair of amino acids)

We can run AlphaFold ourselves if we're not happy with AFDB (i.e. no coverage or poor model)

Interpreting/ Analyzing AF results in R

```
results_dir <- "HIVPR_dimer_23119/"  
results_dir
```

```
[1] "HIVPR_dimer_23119/"
```

The labsheet is very useful!