Class10

Quarto

From last class, we sae that the PDB had 244,290 proteins (Oct 2025). UniprotKB has 199,579,901 total protein sequences.

```
244290/199579901*100
```

```
[1] 0.1224021
```

AFDB is attempting to address this gap.

There are two "Quality Scores" from AlphaFold, one for residues (i.e. each amino acid) called the pLDDT score, and one for confidence in the relative position of the two residues, called PAE.

##Generating your own structure predictions

![]~Downloads/HIVPR_DIMER

Upload picture of 5 generated HIV-PR models

and the top 2 (First and last)

Upload picture of HIV Dimer with uncertainty coloring (plDDT score for model 1 and for model 5)

Custom analysis of resulting models in R

Read key results files into R. The first thing I need to know is what my results directory/folder is called (i.e. the mame is different for every AlhpaFold run/job)

results.dir <- HIVPrDimer_23119_0 results.dir

plot(m1*atom*b[m1$calpha], typ="l", ylim=c(0,100))

##Residue conservation from alignment file

aln_file <- list.files(path=results_dir, pattern=".a3m$", full.names = TRUE) aln_file

aln <- read.fasta(aln_file[1], to.upper = TRUE)

sim <- conserv(aln)

plotb3(sim[1:99], ylab="Conservation Score")

con <- consensus(aln, cutoff = 0.9) con$seq