Class 11 - Alpha Fold

Vivian Chau (A16913056)

Here we analyze our AlphaFold structure prediction models The input directory/folder comes from the ColabFold server: RBP4_0afab.result.zip

To visualize my model PDB files, Mol* was used:

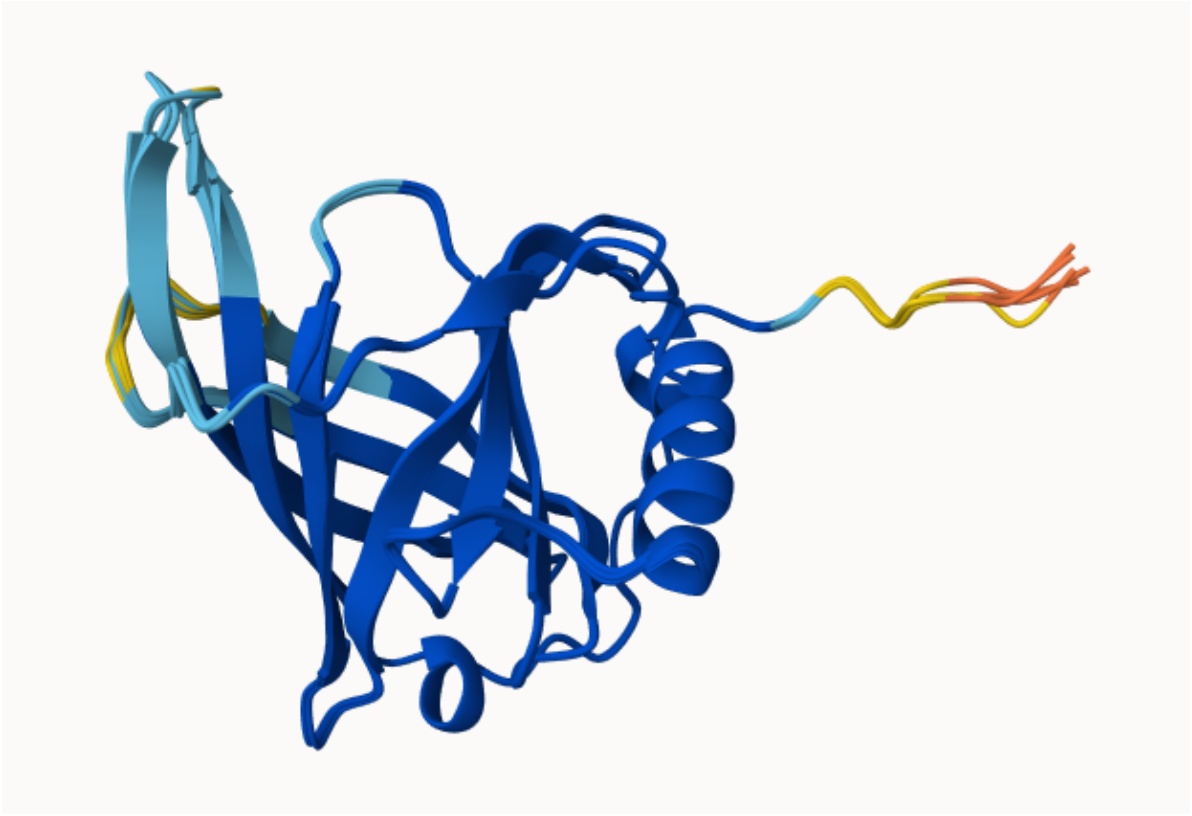


Figure 1: Superposed Model of PDB Structure Colored by pLDDT Scores

```
results_dir <- "RBP4_0afab"
results_dir
```

```
[1] "RBP4_0afab"
```

```
# File names for all PDB models
pdb_files <- list.files(path=results_dir,
                        pattern=".pdb",
                        full.names = TRUE)

# Print our PDB file names
basename(pdb_files)
```

```
[1] "RBP4_0afab_unrelaxed_rank_001_alphafold2_ptm_model_3_seed_000.pdb"
[2] "RBP4_0afab_unrelaxed_rank_002_alphafold2_ptm_model_5_seed_000.pdb"
[3] "RBP4_0afab_unrelaxed_rank_003_alphafold2_ptm_model_4_seed_000.pdb"
[4] "RBP4_0afab_unrelaxed_rank_004_alphafold2_ptm_model_2_seed_000.pdb"
[5] "RBP4_0afab_unrelaxed_rank_005_alphafold2_ptm_model_1_seed_000.pdb"
```

I will use the Bio3D package for analysis

```
library(bio3d)

# Read all data from Models
#  and superpose/fit coords
pdbs <- pdbaln(pdb_files, fit=TRUE, exefile="msa")
```

```
Reading PDB files:
RBP4_0afab/RBP4_0afab_unrelaxed_rank_001_alphafold2_ptm_model_3_seed_000.pdb
RBP4_0afab/RBP4_0afab_unrelaxed_rank_002_alphafold2_ptm_model_5_seed_000.pdb
RBP4_0afab/RBP4_0afab_unrelaxed_rank_003_alphafold2_ptm_model_4_seed_000.pdb
RBP4_0afab/RBP4_0afab_unrelaxed_rank_004_alphafold2_ptm_model_2_seed_000.pdb
RBP4_0afab/RBP4_0afab_unrelaxed_rank_005_alphafold2_ptm_model_1_seed_000.pdb
.....

Extracting sequences

pdb/seq: 1   name: RBP4_0afab/RBP4_0afab_unrelaxed_rank_001_alphafold2_ptm_model_3_seed_000.p
pdb/seq: 2   name: RBP4_0afab/RBP4_0afab_unrelaxed_rank_002_alphafold2_ptm_model_5_seed_000.p
pdb/seq: 3   name: RBP4_0afab/RBP4_0afab_unrelaxed_rank_003_alphafold2_ptm_model_4_seed_000.p
```

```
pdb/seq: 4   name: RBP4_0afab/RBP4_0afab_unrelaxed_rank_004_alphafold2_ptm_model_2_seed_000.p
pdb/seq: 5   name: RBP4_0afab/RBP4_0afab_unrelaxed_rank_005_alphafold2_ptm_model_1_seed_000.p
```

A quick view of these model sequences:

```
pdbs
```

```
                              1        .         .         .         .         50
[Truncated_Name:1]RBP4_0afab   GSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVAEF
[Truncated_Name:2]RBP4_0afab   GSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVAEF
[Truncated_Name:3]RBP4_0afab   GSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVAEF
[Truncated_Name:4]RBP4_0afab   GSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVAEF
[Truncated_Name:5]RBP4_0afab   GSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVAEF
                               **************************************************
                              1        .         .         .         .         50

                             51        .         .         .         .         100
[Truncated_Name:1]RBP4_0afab   SVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVAS
[Truncated_Name:2]RBP4_0afab   SVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVAS
[Truncated_Name:3]RBP4_0afab   SVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVAS
[Truncated_Name:4]RBP4_0afab   SVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVAS
[Truncated_Name:5]RBP4_0afab   SVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVAS
                               **************************************************
                             51        .         .         .         .         100

                            101        .         .         .         .         150
[Truncated_Name:1]RBP4_0afab   FLQKGNDDHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLP
[Truncated_Name:2]RBP4_0afab   FLQKGNDDHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLP
[Truncated_Name:3]RBP4_0afab   FLQKGNDDHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLP
[Truncated_Name:4]RBP4_0afab   FLQKGNDDHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLP
[Truncated_Name:5]RBP4_0afab   FLQKGNDDHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLP
                               **************************************************
                            101        .         .         .         .         150

                            151        .         .        179
[Truncated_Name:1]RBP4_0afab   PEAQKIVRQRQEELCLARQYRLIVHNGYC
[Truncated_Name:2]RBP4_0afab   PEAQKIVRQRQEELCLARQYRLIVHNGYC
[Truncated_Name:3]RBP4_0afab   PEAQKIVRQRQEELCLARQYRLIVHNGYC
[Truncated_Name:4]RBP4_0afab   PEAQKIVRQRQEELCLARQYRLIVHNGYC
[Truncated_Name:5]RBP4_0afab   PEAQKIVRQRQEELCLARQYRLIVHNGYC
                               *****************************
                            151        .         .        179
```

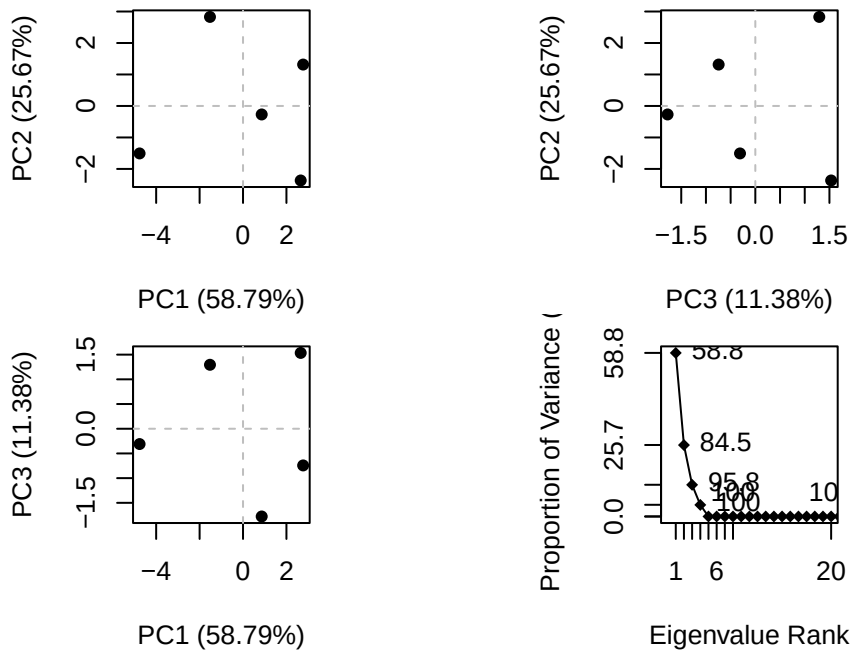
```
Call:
  pdbaln(files = pdb_files, fit = TRUE, exefile = "msa")

Class:
  pdbs, fasta

Alignment dimensions:
  5 sequence rows; 179 position columns (179 non-gap, 0 gap)

+ attr: xyz, resno, b, chain, id, ali, resid, sse, call
```

A Quick PCA

```
pc.rbp<-pca(pdbs)
pc.rbp
```

```
Call:
  pca.pdbs(pdbs = pdbs)

Class:
  pca

Number of eigenvalues:
  537

      Eigenvalue Variance Cumulative
PC 1      10.120   58.792     58.792
PC 2       4.419   25.673     84.464
PC 3       1.959   11.381     95.846
PC 4       0.715    4.154    100.000
PC 5       0.000    0.000    100.000
PC 6       0.000    0.000    100.000

   (Obtained from 5 conformers with 537 xyz input values).

+ attr: L, U, z, au, sdev, mean, call
```

```
plot(pc.rbp)
```



RMSD Analysis

RMSD is a common measure of structural distance used in structural biology

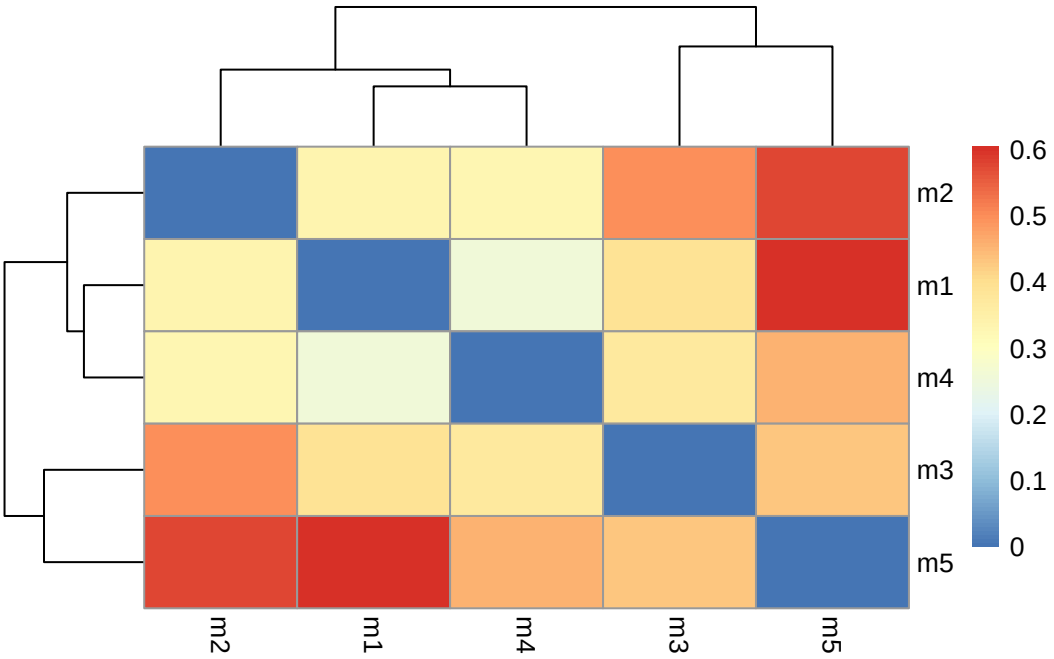
```
rd <- rmsd(pdbs, fit=T)
```

```
Warning in rmsd(pdbs, fit = T): No indices provided, using the 179 non NA positions
```

```
range(rd)
```

```
[1] 0.000 0.604
```

```
library(pheatmap)

colnames(rd) <- paste0("m",1:5)
rownames(rd) <- paste0("m",1:5)
pheatmap(rd)
```


m2
m1
m4
m3
m5
m2
m1
m4
m3
m5
0.6
0.5
0.4
0.3
0.2
0.1
0