AlphaFold Analysis

Bayah Essayem (A17303992)

2025-02-11

Table of contents

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

Change this for YOUR result dir name

```
results_dir <- "hivpr_monomer_94b5b.result/"
```

File names for all PDB models

```
  pdb_files <- list.files(path="hivpr_monomer_94b5b.result/hivpr_monomer_94b5b/",
                          pattern="*.pdb",
                          full.names = TRUE)


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

```
[1] "hivpr_monomer_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000.pdb"
[2] "hivpr_monomer_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000.pdb"
[3] "hivpr_monomer_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000.pdb"
[4] "hivpr_monomer_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000.pdb"
[5] "hivpr_monomer_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000.pdb"
```

I will use the Bio3D package for analysis

```
library(bio3d)
```

Align and superpose

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

```
Reading PDB files:
hivpr_monomer_94b5b.result/hivpr_monomer_94b5b/hivpr_monomer_94b5b_unrelaxed_rank_001_alphafo
hivpr_monomer_94b5b.result/hivpr_monomer_94b5b/hivpr_monomer_94b5b_unrelaxed_rank_002_alphafo
hivpr_monomer_94b5b.result/hivpr_monomer_94b5b/hivpr_monomer_94b5b_unrelaxed_rank_003_alphafo
hivpr_monomer_94b5b.result/hivpr_monomer_94b5b/hivpr_monomer_94b5b_unrelaxed_rank_004_alphafo
hivpr_monomer_94b5b.result/hivpr_monomer_94b5b/hivpr_monomer_94b5b_unrelaxed_rank_005_alphafo
.....

Extracting sequences

pdb/seq: 1   name: hivpr_monomer_94b5b.result/hivpr_monomer_94b5b/hivpr_monomer_94b5b_unrelax
pdb/seq: 2   name: hivpr_monomer_94b5b.result/hivpr_monomer_94b5b/hivpr_monomer_94b5b_unrelax
pdb/seq: 3   name: hivpr_monomer_94b5b.result/hivpr_monomer_94b5b/hivpr_monomer_94b5b_unrelax
pdb/seq: 4   name: hivpr_monomer_94b5b.result/hivpr_monomer_94b5b/hivpr_monomer_94b5b_unrelax
pdb/seq: 5   name: hivpr_monomer_94b5b.result/hivpr_monomer_94b5b/hivpr_monomer_94b5b_unrelax
```

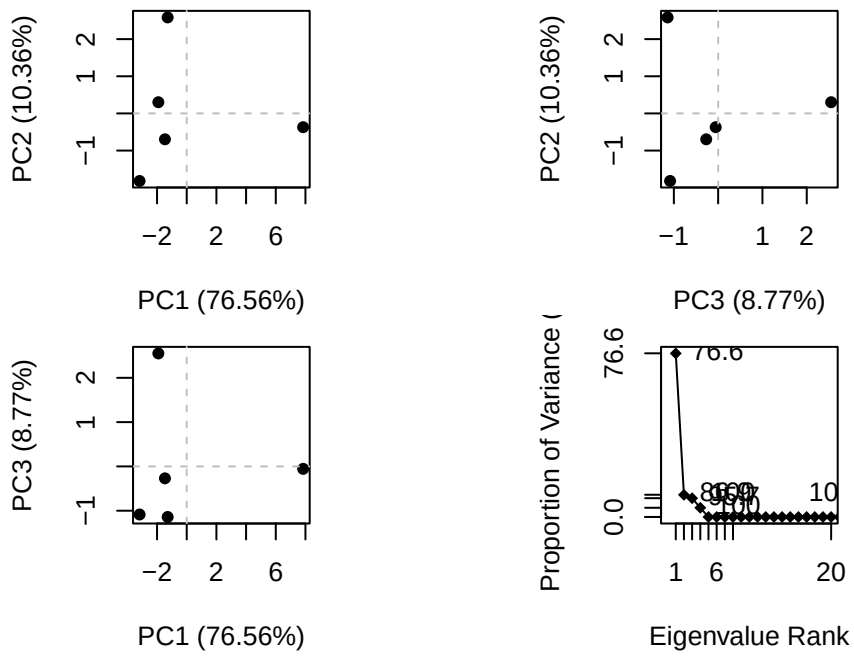
```
pdbs
```

```
                                1        .        .        .        .        50
[Truncated_Name:1]hivpr_mono   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:2]hivpr_mono   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:3]hivpr_mono   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:4]hivpr_mono   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:5]hivpr_mono   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
                               **************************************************
```

```
                                 1        .         .         .         .         50

                                51        .         .         .         .        99
[Truncated_Name:1]hivpr_mono     GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:2]hivpr_mono     GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:3]hivpr_mono     GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:4]hivpr_mono     GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:5]hivpr_mono     GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
                                 *************************************************
                                51        .         .         .         .        99

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

Class:
  pdbs, fasta

Alignment dimensions:
  5 sequence rows; 99 position columns (99 non-gap, 0 gap)

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

A quick PCA

```
pc <- pca(pdbs)
plot(pc)
```



RMSD analysis

RMSD (root, mean, square, of the distance) is a common measure of structural distance used in structural biology.

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

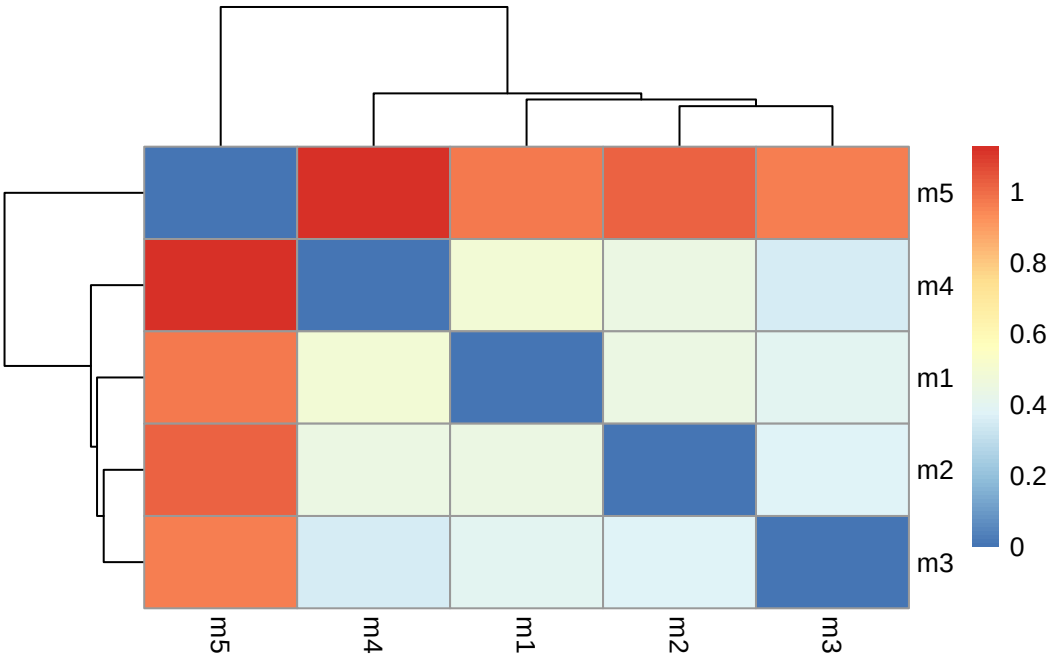
```
Warning in rmsd(pdbs, fit = T): No indices provided, using the 99 non NA positions
```

```
rd
```

```
                                                                    hivpr_monomer_94b5b_u
hivpr_monomer_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000
                                                                    hivpr_monomer_94b5b_u
hivpr_monomer_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000
```

```
hivpr_monomer_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000
                                                                        hivpr_monomer_94b5b_ur
hivpr_monomer_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000
                                                                        hivpr_monomer_94b5b_ur
hivpr_monomer_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000
                                                                        hivpr_monomer_94b5b_ur
hivpr_monomer_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000
hivpr_monomer_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000
```

```r
library(pheatmap)

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


m5
m4
m1
m2
m3
1
0.8
0.6
0.4
0.2
0
m5
m4
m1
m2
m3