

# AlphaFold Analysis

Nhi (PID: A18053310)

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

```
# Change this for YOUR results dir name
results_dir <- "hivpr_monomer_94b5b/"
```

```
# 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] "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

```
pdbbs <- pdbaln(pdb_files, fit=TRUE, exefile="msa")
```

Reading PDB files:

```
hivpr_monomer_94b5b///hivpr_monomer_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000
hivpr_monomer_94b5b///hivpr_monomer_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000
hivpr_monomer_94b5b///hivpr_monomer_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000
hivpr_monomer_94b5b///hivpr_monomer_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000
hivpr_monomer_94b5b///hivpr_monomer_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000
.....
```

Extracting sequences

```
pdb/seq: 1   name: hivpr_monomer_94b5b///hivpr_monomer_94b5b_unrelaxed_rank_001_alphafold2_p
pdb/seq: 2   name: hivpr_monomer_94b5b///hivpr_monomer_94b5b_unrelaxed_rank_002_alphafold2_p
pdb/seq: 3   name: hivpr_monomer_94b5b///hivpr_monomer_94b5b_unrelaxed_rank_003_alphafold2_p
pdb/seq: 4   name: hivpr_monomer_94b5b///hivpr_monomer_94b5b_unrelaxed_rank_004_alphafold2_p
pdb/seq: 5   name: hivpr_monomer_94b5b///hivpr_monomer_94b5b_unrelaxed_rank_005_alphafold2_p
```

pdbs

```

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

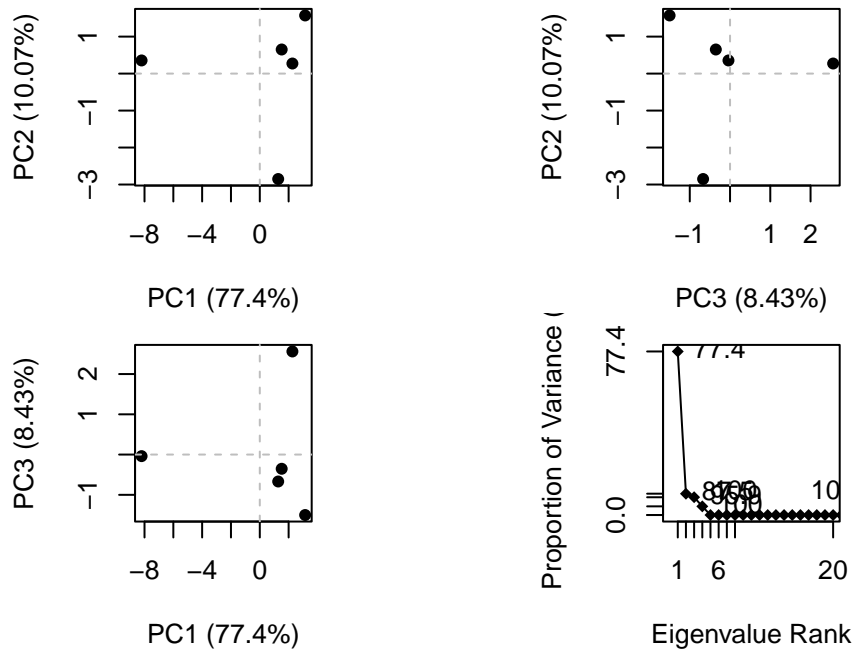
```
pdbs, fasta
```

Alignment dimensions:

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

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

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



## RMSD analysis

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

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

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

```
rd
```

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\_un

```
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\_un

```
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\_un

```
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\_un

```
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
```

```
library(pheatmap)
```

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

