

Class 12 AlphaFold analysis

Nancy Leon Rivera

Here we demonstrate how to analyze and make sense of models from AlphaFold. We begin by reading all the model PDB files...

```
library(bio3d)
```

PDB file names of my models

```
files <- list.files("hiv_monomer_94b5b/", pattern = ".pdb", full.names = T)
files
```

```
[1] "hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000.pdb"
[2] "hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000.pdb"
[3] "hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000.pdb"
[4] "hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000.pdb"
[5] "hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000.pdb"
```

#Align and superimpose

```
pdba <- pdbaln(files, fit = TRUE, exefile="msa")
```

Reading PDB files:

```
hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000.pdb
hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000.pdb
hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000.pdb
hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000.pdb
hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000.pdb
.....
```

Extracting sequences

```
pdb/seq: 1    name: hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_001_alphafold2_ptm_mod
pdb/seq: 2    name: hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_002_alphafold2_ptm_mod
pdb/seq: 3    name: hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_003_alphafold2_ptm_mod
pdb/seq: 4    name: hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_004_alphafold2_ptm_mod
pdb/seq: 5    name: hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_005_alphafold2_ptm_mod
```

RMSD analysis

```
rd <- rmsd(pdb)
```

Warning in rmsd(pdb): No indices provided, using the 99 non NA positions

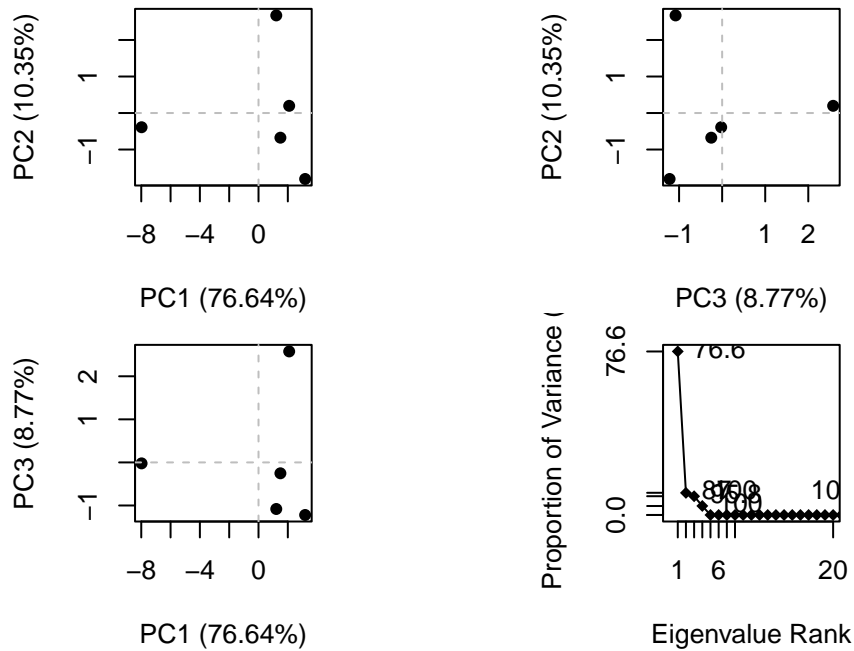
```
mean(rd)
```

```
[1] 0.53448
```

```
source("http://tinyurl.com/newviewngl")
library(NGLViewer)
#view.pdb(pdb)
```

##PCA

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



```
#Residues conservation from alignment file
```

AlphaFold writes out the MSA it calculated

```
aln_file <- list.files(path="hiv_monomer_94b5b/", pattern = ".a3m$", full.names = TRUE)
aln_file
```

```
[1] "hiv_monomer_94b5b/hiv_monomer_94b5b.a3m"
```

```
aln <- read.fasta(aln_file, to.upper=T)
```

```
[1] " ** Duplicated sequence id's: 101 **"
```

```
dim(aln$ali)
```

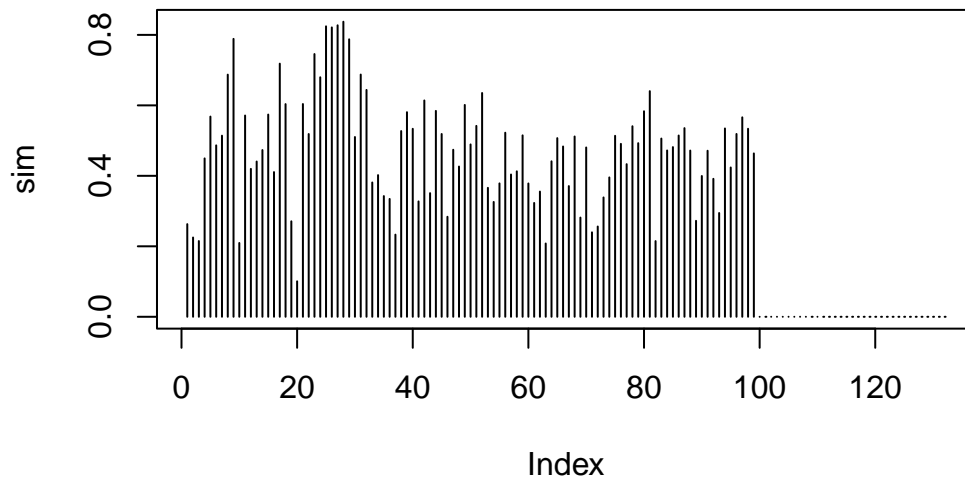
```
[1] 5378 132
```

```
sim <- conserv(aln)
```

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

```
[1] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[19] "-" "-" "-" "-" "-" "-" "D" "T" "G" "A" "-" "-" "-" "-" "-" "-" "-" "-"
[37] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[55] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[73] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[91] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[109] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[127] "-" "-" "-" "-" "-" "-"
```

```
plot(sim, type = "h")
```



```
#pdb <- read.pdb( files[1])
#view.pdb(pdb, backgroundColor = "pink", highlight = atom.select(pdb, resno=25:28), highlight
```