

# AlphaFold Analysis

AUTHOR

Katie Mostoller A17259578

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

```
library(bio3d)
```

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

Align and superpose

```
pdb<- 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_model_5_seed_000.pdb
pdb/seq: 2  name:
hiv_monomer_94b5b//hiv_monomer_94b5b_unrelaxed_rank_002_alphaFold2_ptm_model_4_seed_000.pdb
pdb/seq: 3  name:
hiv_monomer_94b5b//hiv_monomer_94b5b_unrelaxed_rank_003_alphaFold2_ptm_model_1_seed_000.pdb
pdb/seq: 4  name:
hiv_monomer_94b5b//hiv_monomer_94b5b_unrelaxed_rank_004_alphaFold2_ptm_model_3_seed_000.pdb
pdb/seq: 5  name:
hiv_monomer_94b5b//hiv_monomer_94b5b_unrelaxed_rank_005_alphaFold2_ptm_model_2_seed_000.pdb
```

RMSD analysis

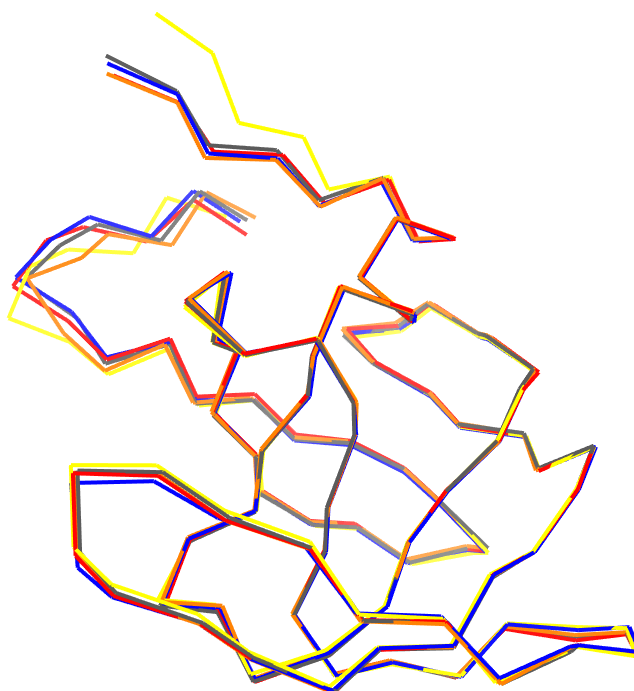
```
rd <- rmsd(pdbbs)
```

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

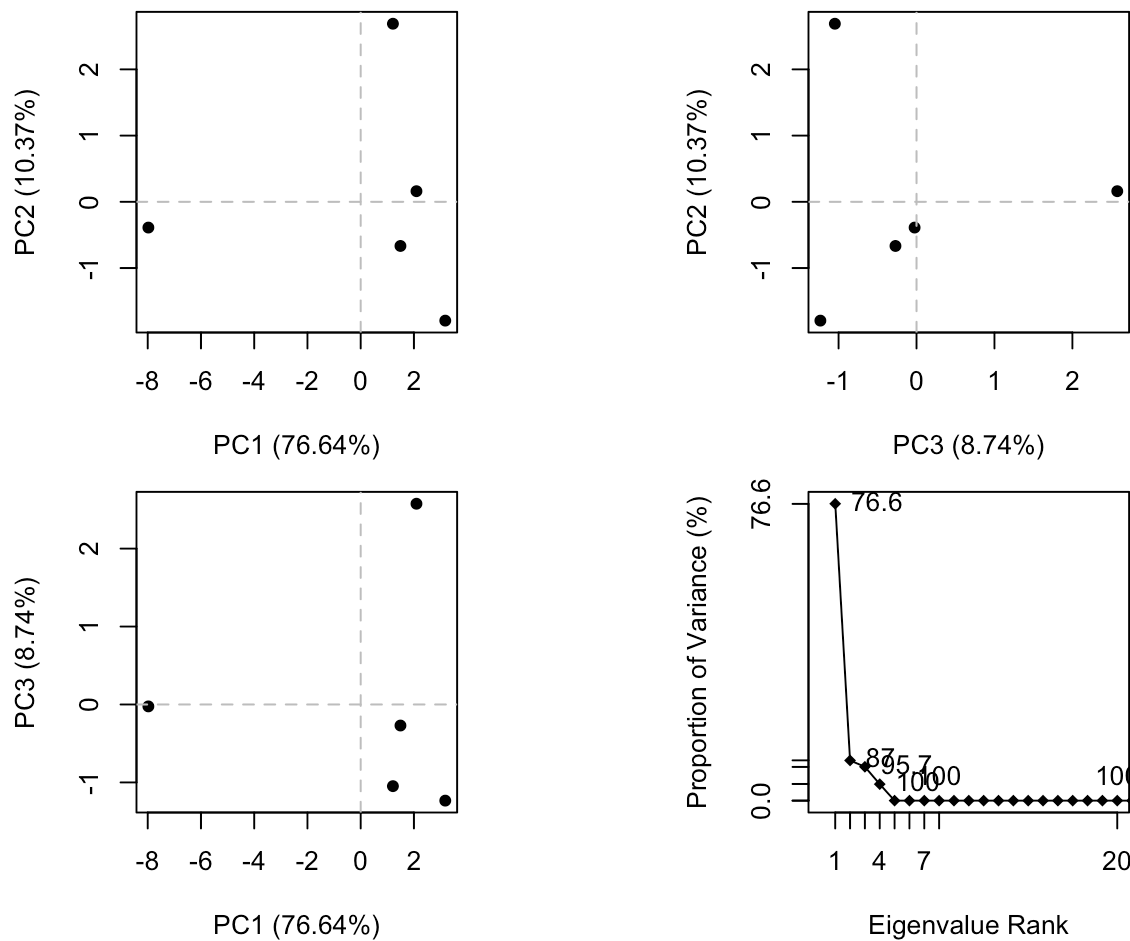
```
mean(rd)
```

```
[1] 0.53432
```

```
source("https://tinyurl.com/newviewngl")  
library(NGLVieweR)  
view.pdbbs(pdbbs)
```



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



## Residue conservation from alignment file

```
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 = TRUE)
```

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

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

```
[1] 5378 132
```

Score residue conservation

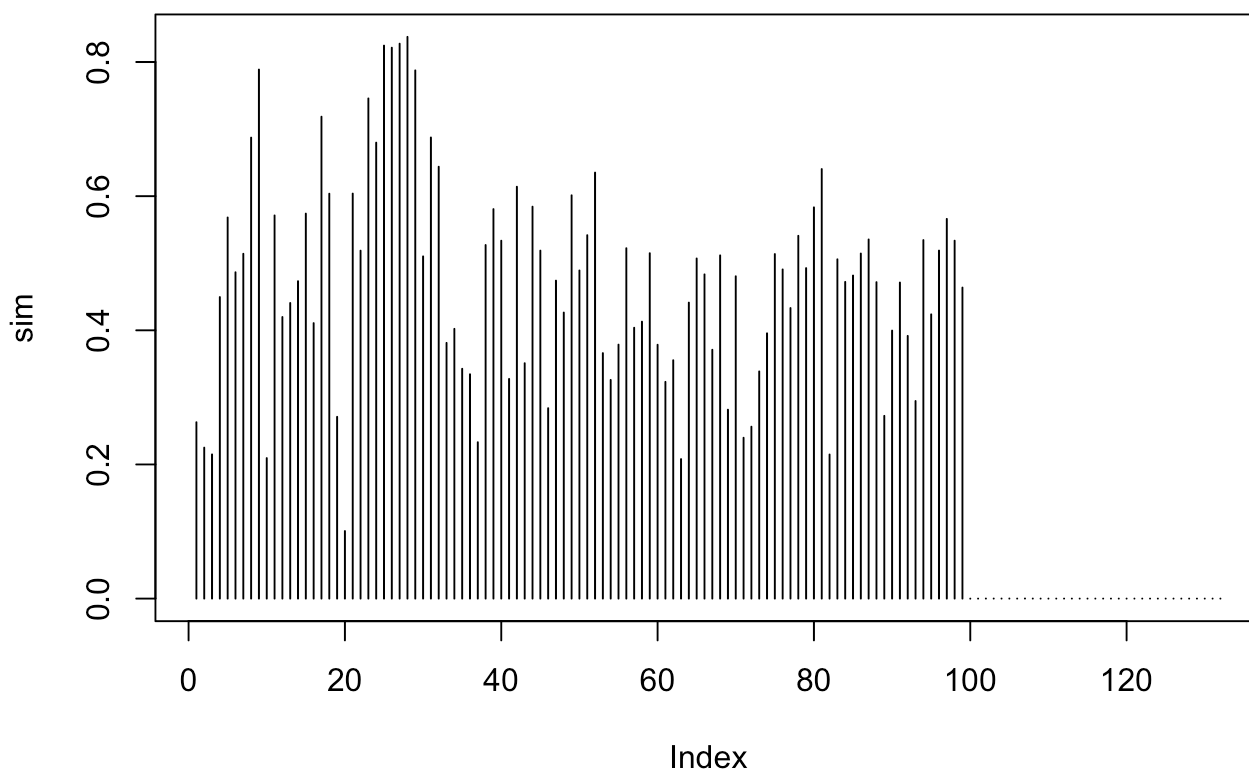
```
sim <- conserv(aln)
```

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

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

```
pdb <- read.pdb(files[1])
```

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



```
pdb <- read.pdb(files[1])

view.pdb(pdb, backgroundColor = "pink",
         highlight = atom.select(pdb, resno = 25:28))
```

