# AlphaFold analysis

Bryn Baxter PID A69038039

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 model

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

```
[1] "hiv_monomer_94b5b.result/hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_001_alphafol
[2] "hiv_monomer_94b5b.result/hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_002_alphafol
[3] "hiv_monomer_94b5b.result/hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_003_alphafol
[4] "hiv_monomer_94b5b.result/hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_004_alphafol
[5] "hiv_monomer_94b5b.result/hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_005_alphafol
```

Align and superpose

```
pdbs <- pdbaln(files, fit=T, exefile="msa")
```

```
Reading PDB files:
hiv_monomer_94b5b.result/hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_001_alphafold2_pt
hiv_monomer_94b5b.result/hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_002_alphafold2_pt
hiv_monomer_94b5b.result/hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_003_alphafold2_pt
hiv_monomer_94b5b.result/hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_004_alphafold2_pt
hiv_monomer_94b5b.result/hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_rank_005_alphafold2_pt
.....
```

```
Extracting sequences

pdb/seq: 1   name: hiv_monomer_94b5b.result/hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_ra
pdb/seq: 2   name: hiv_monomer_94b5b.result/hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_ra
pdb/seq: 3   name: hiv_monomer_94b5b.result/hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_ra
pdb/seq: 4   name: hiv_monomer_94b5b.result/hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_ra
pdb/seq: 5   name: hiv_monomer_94b5b.result/hiv_monomer_94b5b/hiv_monomer_94b5b_unrelaxed_ra
```

RMSD analysis

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

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

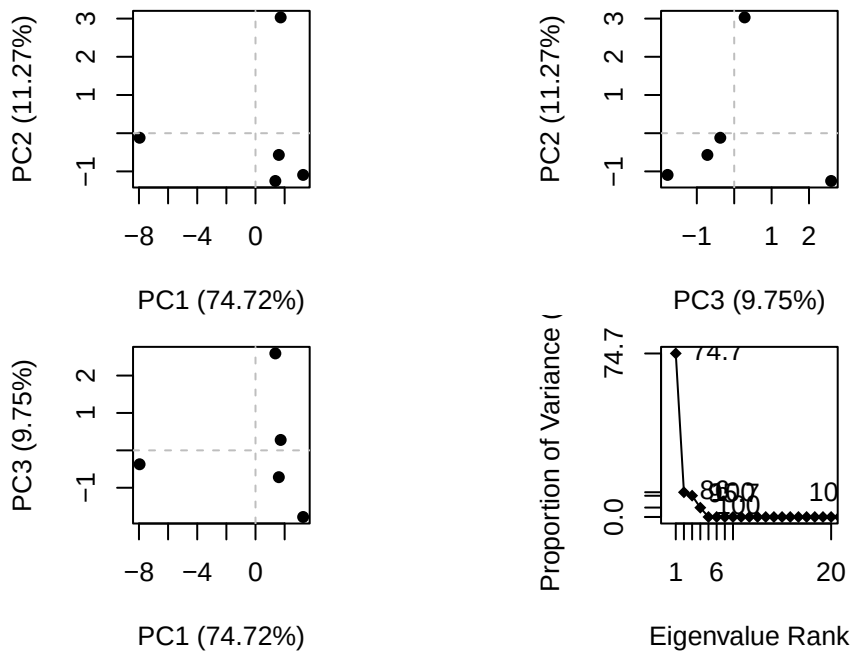
```
mean(rd)
```

```
[1] 0.54392
```

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

## PCA

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




#Residue conservation from alignment file

Alphafold

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

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

Score residue conservation:

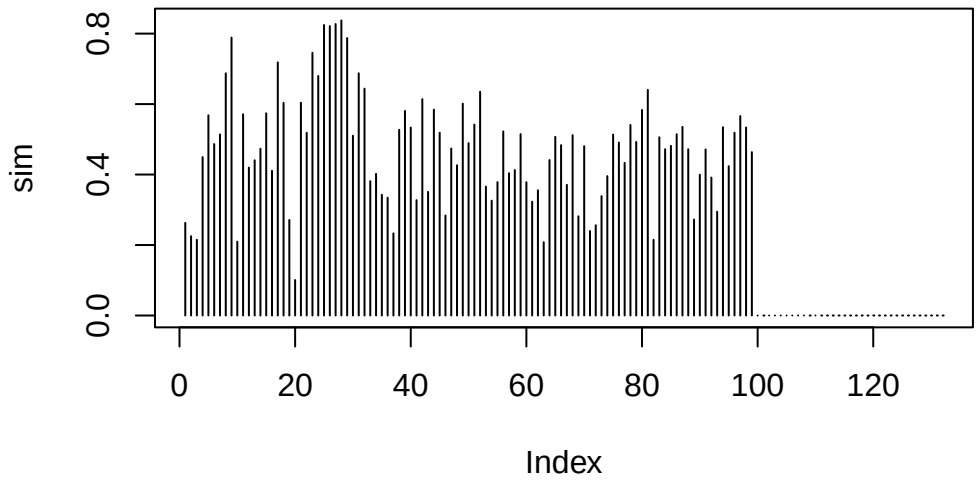
```
sim <- conserv(aln)
```

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

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

PLot the conservation along the sequence/structure

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




Lets look at these conserved positions in the structure:

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

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