

# lab11 - AlphaFold

AUTHOR

Karina Cardenas, A16742606

## Generating your own structure predictions

Here we read the results from AlphaFold and try to interpret all the models and quality score metrics.

```
library(bio3d)

pth <- "dimer_23119/"

pdb.files <- list.files(path = pth, full.names = TRUE, pattern = ".pdb")
```

```
file.exists(pdb.files)
```

```
[1] TRUE TRUE TRUE TRUE TRUE
```

```
pdbx <- pdbaln(pdb.files, fit = TRUE, exefile="msa")
```

### Reading PDB files:

```
dimer_23119//dimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_2_seed_000.pdb
dimer_23119//dimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000.pdb
dimer_23119//dimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000.pdb
dimer_23119//dimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_1_seed_000.pdb
dimer_23119//dimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000.pdb
.....
```

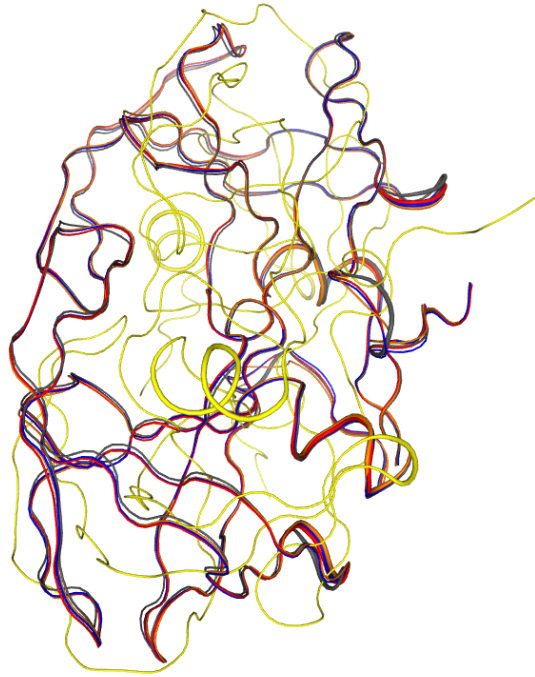
## Extracting sequences

```

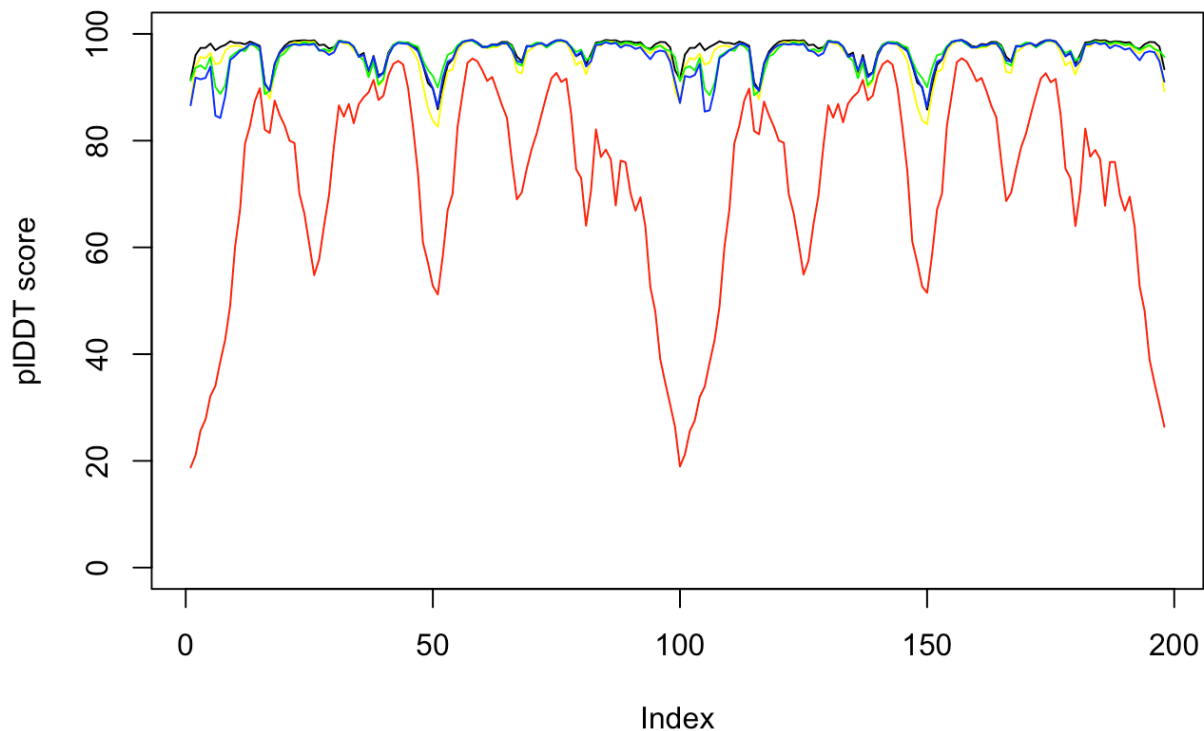
pdb/seq: 1    name:
dimer_23119//dimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_2_seed_000.pdb
pdb/seq: 2    name:
dimer_23119//dimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000.pdb
pdb/seq: 3    name:
dimer_23119//dimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000.pdb
pdb/seq: 4    name:
dimer_23119//dimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_1_seed_000.pdb
pdb/seq: 5    name:
dimer_23119//dimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000.pdb

```

```
library(bio3dview)
view.pdbs(pdbs)
```



```
plot(pdbb$b[1,], typ = "l", ylim = c(0,100), ylab = "pLDDT score")  
  
lines(pdbb$b[2,], typ = "l", col = "yellow")  
  
lines(pdbb$b[3,], typ = "l", col = "green")  
  
lines(pdbb$b[4,], typ = "l", col = "blue")  
  
lines(pdbb$b[5,], typ = "l", col = "red")
```



Alpha fold returns it's large alignment file used for analysis. Here we read this file and score conversation per position.

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

```
[1] "dimer_23119//dimer_23119.a3m"
```

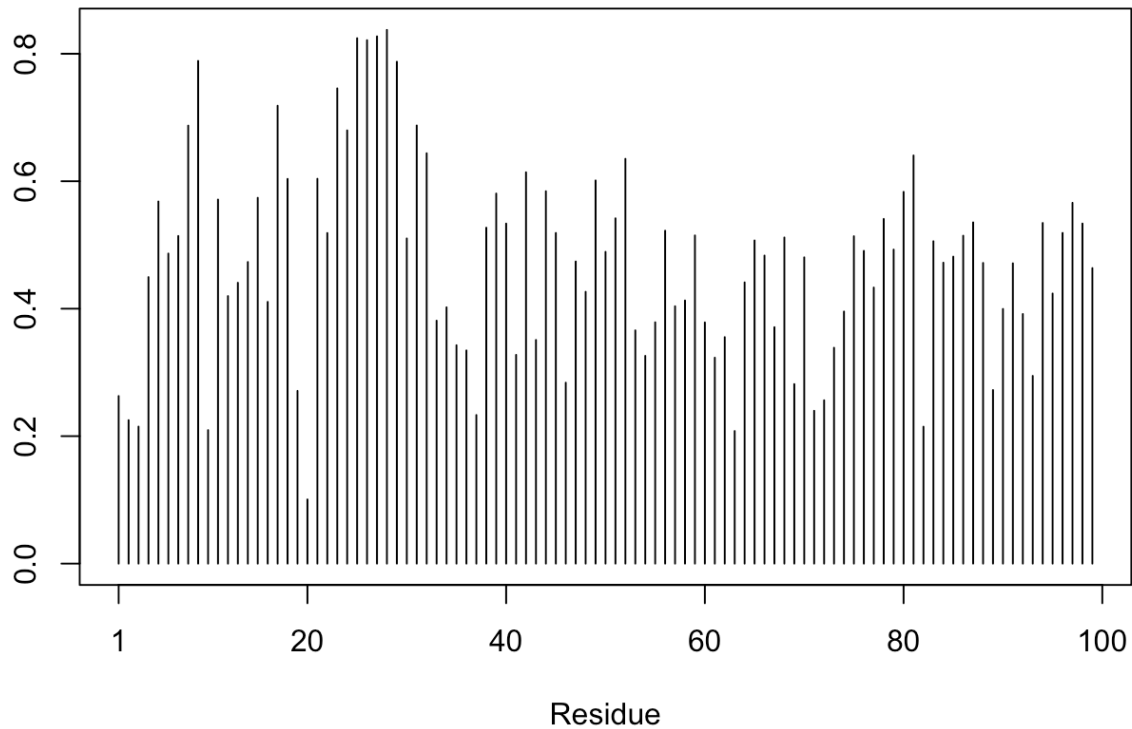
```
aln <- read.fasta(aln_file[1], to.upper = TRUE)
```

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

We can score residue conservation in the alignment with the `conserv()` function.

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

```
plotb3(sim[1:99])
```



```
ylab="Conservation Score"
```

Find the consensus sequence

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

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

## Predicted alignment Error for Domains

```
library(jsonlite)

# Listing of all PAE JSON files
pae_files <- list.files(path=pth,
                        pattern=".*model.*\\.json",
                        full.names = TRUE)
```

```

pae1 <- read_json(pae_files[1],simplifyVector = TRUE)
pae5 <- read_json(pae_files[5],simplifyVector = TRUE)

attributes(pae1)

```

```

$names
[1] "plddt"  "max_pae" "pae"     "ptm"     "iptm"

```

```

head(pae1$plddt)

```

```

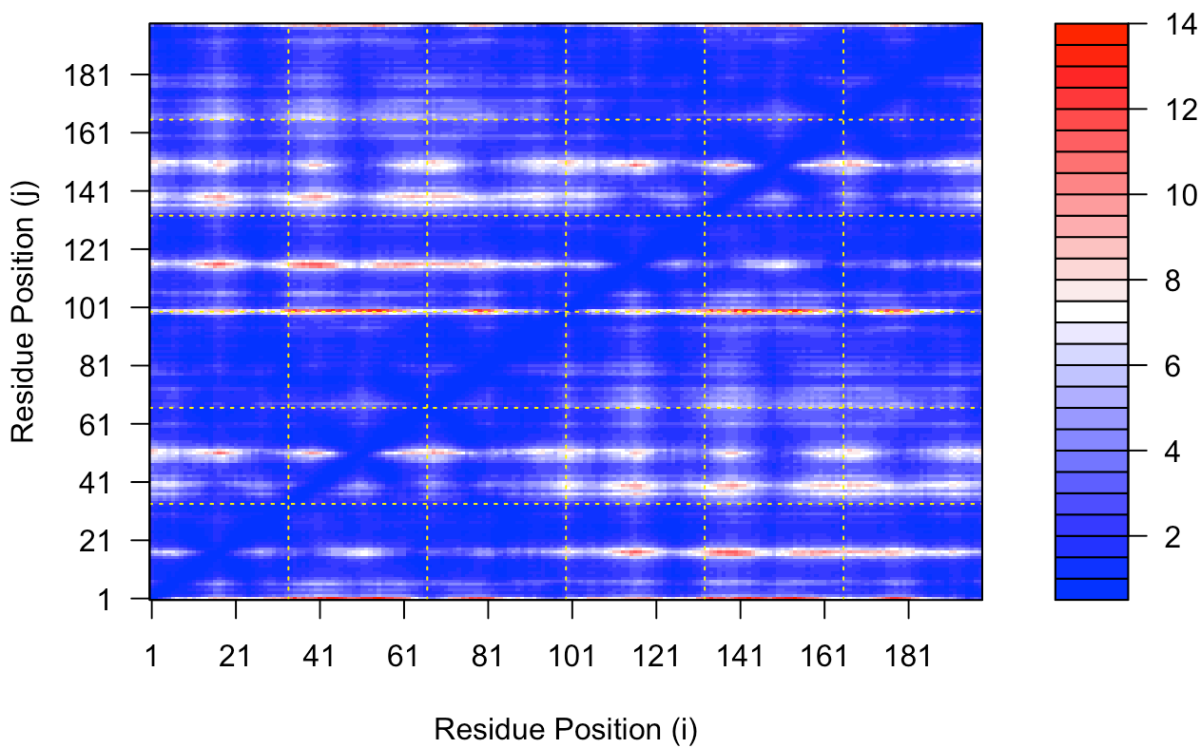
[1] 91.44 96.06 97.38 97.38 98.19 96.94

```

```

plot.dmat(pae1$pae,
          xlab="Residue Position (i)",
          ylab="Residue Position (j)")

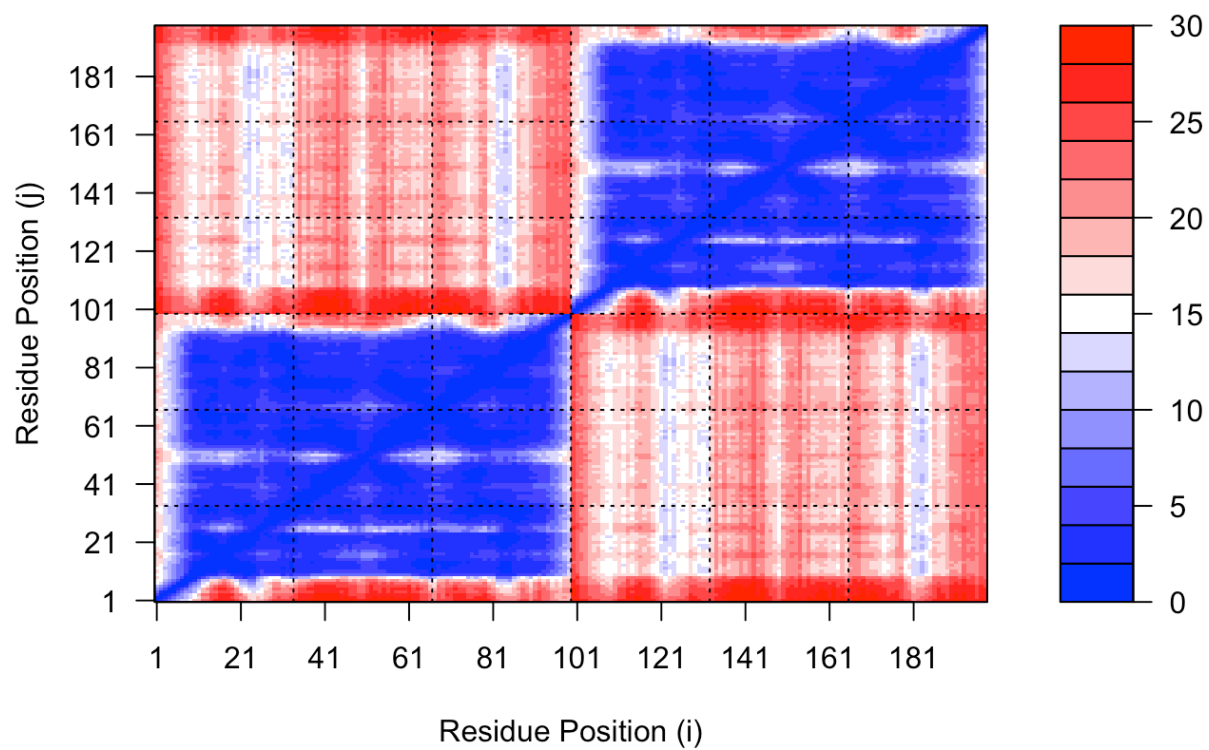
```



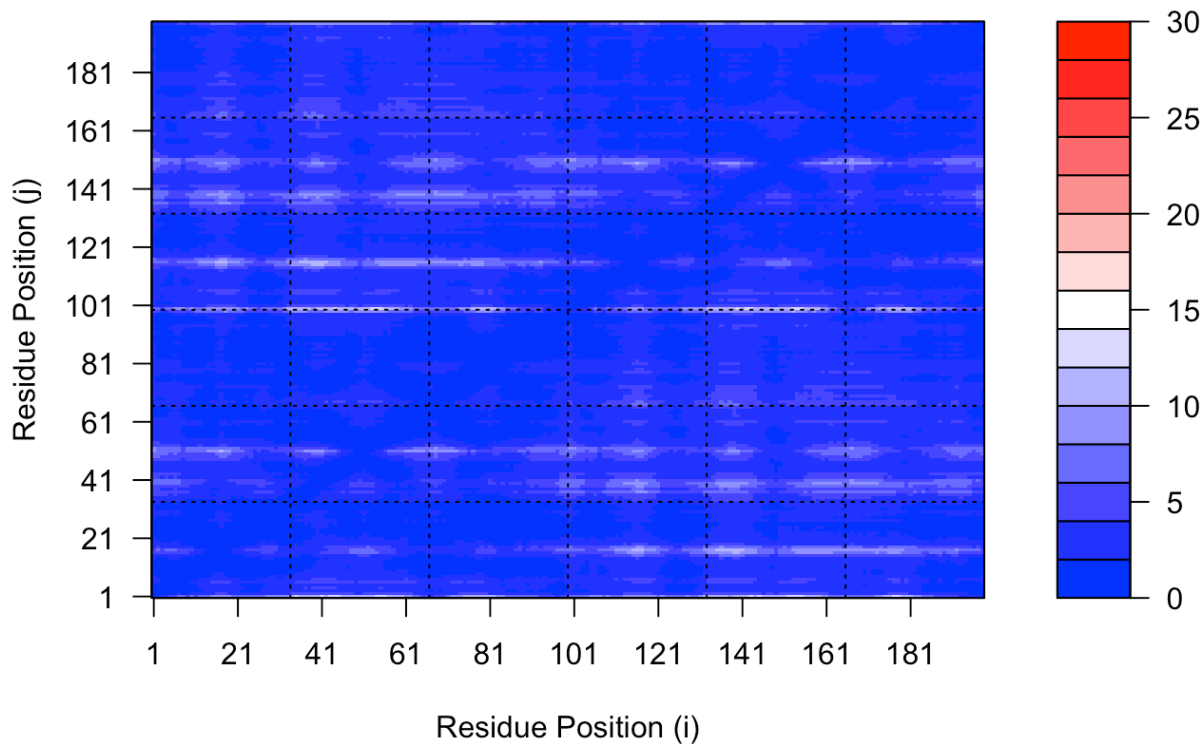
```

plot.dmat(pae5$pae,
          xlab="Residue Position (i)",
          ylab="Residue Position (j)",
          grid.col = "black",
          zlim=c(0,30))

```



```
plot.dmat(pae1$paes,
  xlab="Residue Position (i)",
  ylab="Residue Position (j)",
  grid.col = "black",
  zlim=c(0,30))
```



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

Warning in rmsd(pdbbs, fit = T): No indices provided, using the 198 non NA positions

```
rd
```

```
dimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_2_seed_000
dimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_2_seed_000
0.000
```

```
dimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
0.157
```

```
dimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
0.367
```

```
dimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_1_seed_000
0.308
```

```
dimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000
13.306
```

```
dimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
dimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_2_seed_000
0.157
```

```
dimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
0.000
```

```
dimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
```

```

0.383
dimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_1_seed_000
0.318
dimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000
13.284

dimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
dimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_2_seed_000
0.367
dimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
0.383
dimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
0.000
dimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_1_seed_000
0.483
dimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000
13.406

dimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_1_seed_000
dimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_2_seed_000
0.308
dimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
0.318
dimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
0.483
dimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_1_seed_000
0.000
dimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000
13.247

dimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000
dimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_2_seed_000
13.306
dimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
13.284
dimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
13.406
dimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_1_seed_000
13.247
dimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000
0.000

```

```
library(pheatmap)
```

```

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

```



