

# Class11: alphafold 2 analysis

Aaron Liu A13908620

```
library(bio3d)
```

Here we post process and inspect our modeling from AlphFold2

My results from AF live in the folder/directory hivprdimer\_23119

```
results_dir<-"hivprdimer_23119/"
pdb_files<-list.files(results_dir, pattern=".pdb", full.names = T)
```

We first need to align and superpose the PDB models and we can use the `pdbsaln` function for this:

```
pdbs<-pdbsaln(pdb_files, fit=T, exefile='msa')
```

Reading PDB files:

```
hivprdimer_23119//hivprdimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_00
hivprdimer_23119//hivprdimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_00
hivprdimer_23119//hivprdimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_00
hivprdimer_23119//hivprdimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_00
hivprdimer_23119//hivprdimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_00
.....
```

Extracting sequences

```
pdb/seq: 1    name: hivprdimer_23119//hivprdimer_23119_unrelaxed_rank_001_alphafold2_multimer
pdb/seq: 2    name: hivprdimer_23119//hivprdimer_23119_unrelaxed_rank_002_alphafold2_multimer
pdb/seq: 3    name: hivprdimer_23119//hivprdimer_23119_unrelaxed_rank_003_alphafold2_multimer
pdb/seq: 4    name: hivprdimer_23119//hivprdimer_23119_unrelaxed_rank_004_alphafold2_multimer
pdb/seq: 5    name: hivprdimer_23119//hivprdimer_23119_unrelaxed_rank_005_alphafold2_multimer
```

##The RMSD matrix

A common measure of structural dis-similarity is called RMSD

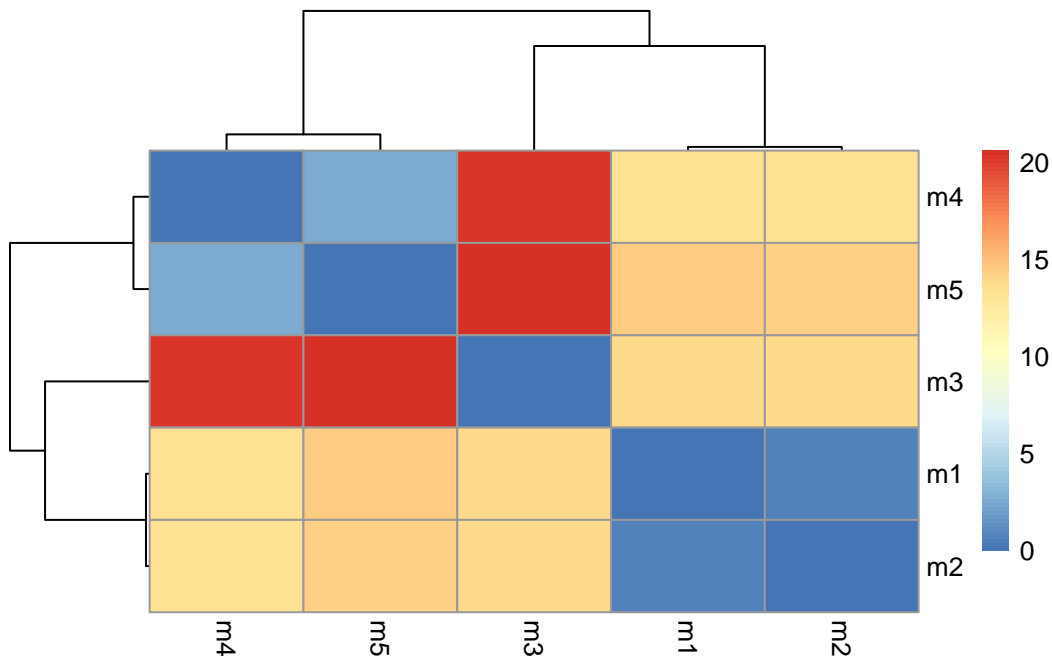
```
rd<-rmsd(pdb)
```

Warning in rmsd(pdb): No indices provided, using the 198 non NA positions

```
rd
```

```
hivprdimer_23119.
hivprdimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000
hivprdimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
hivprdimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
hivprdimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_000
hivprdimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000
hivprdimer_23119.
hivprdimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000
hivprdimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
hivprdimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
hivprdimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_000
hivprdimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000
hivprdimer_23119.
hivprdimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000
hivprdimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
hivprdimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
hivprdimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_000
hivprdimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000
hivprdimer_23119.
hivprdimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000
hivprdimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
hivprdimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
hivprdimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_000
hivprdimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000
hivprdimer_23119.
hivprdimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000
hivprdimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
hivprdimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
hivprdimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_000
hivprdimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000
```

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



Let's view these in Mol\*. Here we want the fitted coords

```
xyz<-pdbfit(pdb, outpath="fitted")
```

Focus on the most invariant part so the fitting and superposition will fit well.

```
core<-core.find(pdb)
```

```
core size 197 of 198 vol = 4705.621
core size 196 of 198 vol = 3968.518
core size 195 of 198 vol = 3714.79
core size 194 of 198 vol = 3483.747
core size 193 of 198 vol = 3301.729
core size 192 of 198 vol = 3100.7
core size 191 of 198 vol = 2947.616
core size 190 of 198 vol = 2824.451
```

core size 189 of 198	vol = 2751.594
core size 188 of 198	vol = 2694.616
core size 187 of 198	vol = 2658.687
core size 186 of 198	vol = 2640.03
core size 185 of 198	vol = 2686.901
core size 184 of 198	vol = 2752.044
core size 183 of 198	vol = 2869.057
core size 182 of 198	vol = 3012.085
core size 181 of 198	vol = 3124.103
core size 180 of 198	vol = 3206.778
core size 179 of 198	vol = 3245.849
core size 178 of 198	vol = 3287.095
core size 177 of 198	vol = 3296.508
core size 176 of 198	vol = 3264.588
core size 175 of 198	vol = 3221.919
core size 174 of 198	vol = 3122.71
core size 173 of 198	vol = 3012.961
core size 172 of 198	vol = 2900.383
core size 171 of 198	vol = 2799.364
core size 170 of 198	vol = 2727.189
core size 169 of 198	vol = 2641.304
core size 168 of 198	vol = 2568.381
core size 167 of 198	vol = 2490.264
core size 166 of 198	vol = 2419.832
core size 165 of 198	vol = 2343.183
core size 164 of 198	vol = 2244.642
core size 163 of 198	vol = 2150.228
core size 162 of 198	vol = 2086.112
core size 161 of 198	vol = 2002.064
core size 160 of 198	vol = 1924.293
core size 159 of 198	vol = 1865.077
core size 158 of 198	vol = 1789.637
core size 157 of 198	vol = 1724.694
core size 156 of 198	vol = 1658.625
core size 155 of 198	vol = 1604.024
core size 154 of 198	vol = 1536.718
core size 153 of 198	vol = 1483.978
core size 152 of 198	vol = 1425.457
core size 151 of 198	vol = 1355.586
core size 150 of 198	vol = 1291.721
core size 149 of 198	vol = 1244.687
core size 148 of 198	vol = 1185.617
core size 147 of 198	vol = 1139.699

core size 146 of 198	vol = 1097.377
core size 145 of 198	vol = 1058.96
core size 144 of 198	vol = 1009.933
core size 143 of 198	vol = 963.52
core size 142 of 198	vol = 914.157
core size 141 of 198	vol = 879.153
core size 140 of 198	vol = 848.54
core size 139 of 198	vol = 805.199
core size 138 of 198	vol = 771.239
core size 137 of 198	vol = 733.808
core size 136 of 198	vol = 692.1
core size 135 of 198	vol = 654.952
core size 134 of 198	vol = 625.318
core size 133 of 198	vol = 595.666
core size 132 of 198	vol = 564.326
core size 131 of 198	vol = 537.596
core size 130 of 198	vol = 516.739
core size 129 of 198	vol = 496.948
core size 128 of 198	vol = 472.202
core size 127 of 198	vol = 445.559
core size 126 of 198	vol = 413.187
core size 125 of 198	vol = 387.299
core size 124 of 198	vol = 367.382
core size 123 of 198	vol = 353.489
core size 122 of 198	vol = 324.965
core size 121 of 198	vol = 301.495
core size 120 of 198	vol = 280.998
core size 119 of 198	vol = 259.747
core size 118 of 198	vol = 246.492
core size 117 of 198	vol = 231.909
core size 116 of 198	vol = 224.119
core size 115 of 198	vol = 213.435
core size 114 of 198	vol = 202.239
core size 113 of 198	vol = 184.344
core size 112 of 198	vol = 166.888
core size 111 of 198	vol = 151.741
core size 110 of 198	vol = 140.588
core size 109 of 198	vol = 127.22
core size 108 of 198	vol = 115.962
core size 107 of 198	vol = 107.685
core size 106 of 198	vol = 100.052
core size 105 of 198	vol = 93.346
core size 104 of 198	vol = 85.241

```

core size 103 of 198  vol = 76.756
core size 102 of 198  vol = 70.999
core size 101 of 198  vol = 65.695
core size 100 of 198  vol = 60.462
core size 99 of 198   vol = 55.377
core size 98 of 198   vol = 50.408
core size 97 of 198   vol = 46.042
core size 96 of 198   vol = 40.63
core size 95 of 198   vol = 33.484
core size 94 of 198   vol = 25.949
core size 93 of 198   vol = 18.967
core size 92 of 198   vol = 13.694
core size 91 of 198   vol = 7.979
core size 90 of 198   vol = 5.169
core size 89 of 198   vol = 3.343
core size 88 of 198   vol = 2.363
core size 87 of 198   vol = 1.836
core size 86 of 198   vol = 1.516
core size 85 of 198   vol = 1.289
core size 84 of 198   vol = 1.109
core size 83 of 198   vol = 0.917
core size 82 of 198   vol = 0.801
core size 81 of 198   vol = 0.68
core size 80 of 198   vol = 0.577
core size 79 of 198   vol = 0.524
core size 78 of 198   vol = 0.481
FINISHED: Min vol ( 0.5 ) reached

```

```

core.inds<-core
xyz<-pdbfit(pdb, inds=core.inds, outpath="core_fitted")

```

To evaluate how good multi-chain or multi-domain models are we need to look at the PAE scores (predicted aligned error)

They are output in the json file

```

pae_files<-list.files(results_dir, pattern="000.json", full.names=T)
pae_files

```

```

[1] "hivprdimer_23119//hivprdimer_23119_scores_rank_001_alphafold2_multimer_v3_model_1_seed_0"
[2] "hivprdimer_23119//hivprdimer_23119_scores_rank_002_alphafold2_multimer_v3_model_5_seed_0"
[3] "hivprdimer_23119//hivprdimer_23119_scores_rank_003_alphafold2_multimer_v3_model_4_seed_0"

```

```
[4] "hivprdimer_23119//hivprdimer_23119_scores_rank_004_alphafold2_multimer_v3_model_2_seed_0"
[5] "hivprdimer_23119//hivprdimer_23119_scores_rank_005_alphafold2_multimer_v3_model_3_seed_0"
```

```
library(jsonlite)
pae1<-read_json(pae_files[1], simplifyVector=T)
attributes(pae1)
```

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

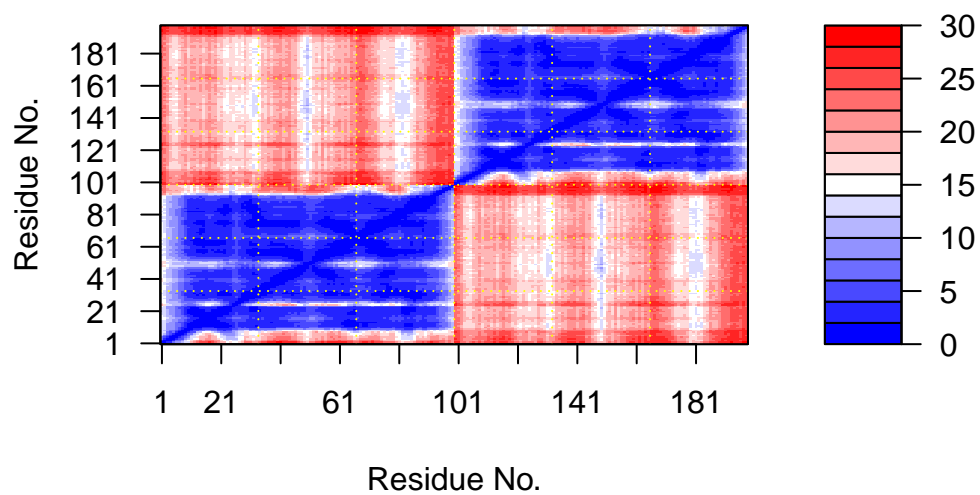
```
pae1$max_pae
```

```
[1] 15.83594
```

```
pae5<-read_json(pae_files[5], simplifyVector=T)
pae5$max_pae
```

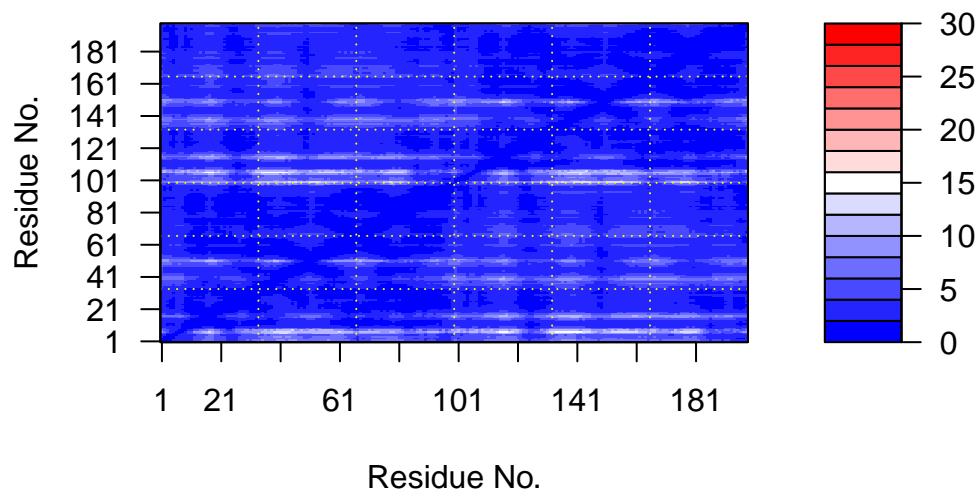
```
[1] 29.25
```

```
plot.dmat(pae5$pae,
          xlab="Residue No.",
          ylab="Residue No.",
          zlim=c(0,30))
```



```
plot.dmat(pae1$pae,
          xlab="Residue No.",
          ylab="Residue No.",
          zlim=c(0,30))
```





##Main point

We can run Alpha fold in google compute We can read these results into R