

AlphaFold Analysis

Ashley Allen (PID: A14633373)

Here we analyze our AlphaFold structure prediction models. The input directory/folder comes from the ColabFold server:

```
#results_dir <- "hivpr_monomer_94b5b"

#pdb_files <- list.files(path=results_dir,
                        #pattern="*.pdb",
                        #full.names = TRUE)

# Print our PDB file names
#basename(pdb_files)
```

I will use the Bio3D package for analysis

```
#library(bio3d)

#pdbs <- pdbaln(pdb_files, fit=TRUE, exefile="msa")
```

```
#pdbs
```

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

#range(rd)
```

```
#library(pheatmap)

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

Find a Gene Project

Analyzing our AlphaFold structure prediction models for my protein.

```
results_dir <- "gene_project_cDNA_f939f"
```

```
pdb_files <- list.files(path=results_dir, pattern="*.pdb", full.names = TRUE)
basename(pdb_files)
```

```
[1] "gene_project_cDNA_f939f_unrelaxed_rank_001_alphafold2_ptm_model_1_seed_000.pdb"
[2] "gene_project_cDNA_f939f_unrelaxed_rank_002_alphafold2_ptm_model_3_seed_000.pdb"
[3] "gene_project_cDNA_f939f_unrelaxed_rank_003_alphafold2_ptm_model_4_seed_000.pdb"
[4] "gene_project_cDNA_f939f_unrelaxed_rank_004_alphafold2_ptm_model_2_seed_000.pdb"
[5] "gene_project_cDNA_f939f_unrelaxed_rank_005_alphafold2_ptm_model_5_seed_000.pdb"
```

```
library(bio3d)
```

Warning: package 'bio3d' was built under R version 4.3.3

```
pdbbs <- pdbaln(pdb_files, fit=TRUE, exefile="msa")
```

Reading PDB files:

```
gene_project_cDNA_f939f/gene_project_cDNA_f939f_unrelaxed_rank_001_alphafold2_ptm_model_1_seed_000.pdb
gene_project_cDNA_f939f/gene_project_cDNA_f939f_unrelaxed_rank_002_alphafold2_ptm_model_3_seed_000.pdb
gene_project_cDNA_f939f/gene_project_cDNA_f939f_unrelaxed_rank_003_alphafold2_ptm_model_4_seed_000.pdb
gene_project_cDNA_f939f/gene_project_cDNA_f939f_unrelaxed_rank_004_alphafold2_ptm_model_2_seed_000.pdb
gene_project_cDNA_f939f/gene_project_cDNA_f939f_unrelaxed_rank_005_alphafold2_ptm_model_5_seed_000.pdb
.....
```

Extracting sequences

```
pdb/seq: 1    name: gene_project_cDNA_f939f/gene_project_cDNA_f939f_unrelaxed_rank_001_alphafold2_ptm_model_1_seed_000.pdb
pdb/seq: 2    name: gene_project_cDNA_f939f/gene_project_cDNA_f939f_unrelaxed_rank_002_alphafold2_ptm_model_3_seed_000.pdb
pdb/seq: 3    name: gene_project_cDNA_f939f/gene_project_cDNA_f939f_unrelaxed_rank_003_alphafold2_ptm_model_4_seed_000.pdb
pdb/seq: 4    name: gene_project_cDNA_f939f/gene_project_cDNA_f939f_unrelaxed_rank_004_alphafold2_ptm_model_2_seed_000.pdb
pdb/seq: 5    name: gene_project_cDNA_f939f/gene_project_cDNA_f939f_unrelaxed_rank_005_alphafold2_ptm_model_5_seed_000.pdb
```

```
pdbbs
```



```

201      .      .      .      .      250

251      .      .      .      .      300
[Truncated_Name:1]gene_proje  REQPPRFAQPGSFHEYAMRWKALIEKQQQDQVDRNIKEAREKLEMEM
[Truncated_Name:2]gene_proje  REQPPRFAQPGSFHEYAMRWKALIEKQQQDQVDRNIKEAREKLEMEM
[Truncated_Name:3]gene_proje  REQPPRFAQPGSFHEYAMRWKALIEKQQQDQVDRNIKEAREKLEMEM
[Truncated_Name:4]gene_proje  REQPPRFAQPGSFHEYAMRWKALIEKQQQDQVDRNIKEAREKLEMEM
[Truncated_Name:5]gene_proje  REQPPRFAQPGSFHEYAMRWKALIEKQQQDQVDRNIKEAREKLEMEM
                                *****
251      .      .      .      .      300

301      .      315
[Truncated_Name:1]gene_proje  EAARHEHQVLMRQD
[Truncated_Name:2]gene_proje  EAARHEHQVLMRQD
[Truncated_Name:3]gene_proje  EAARHEHQVLMRQD
[Truncated_Name:4]gene_proje  EAARHEHQVLMRQD
[Truncated_Name:5]gene_proje  EAARHEHQVLMRQD
                                *****
301      .      315

```

Call:

```
pdaln(files = pdb_files, fit = TRUE, exe_file = "msa")
```

Class:

```
pdbs, fasta
```

Alignment dimensions:

```
5 sequence rows; 315 position columns (315 non-gap, 0 gap)
```

```
+ attr: xyz, resno, b, chain, id, ali, resid, sse, call
```

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

Warning in rmsd(pdb, fit = T): No indices provided, using the 315 non NA positions

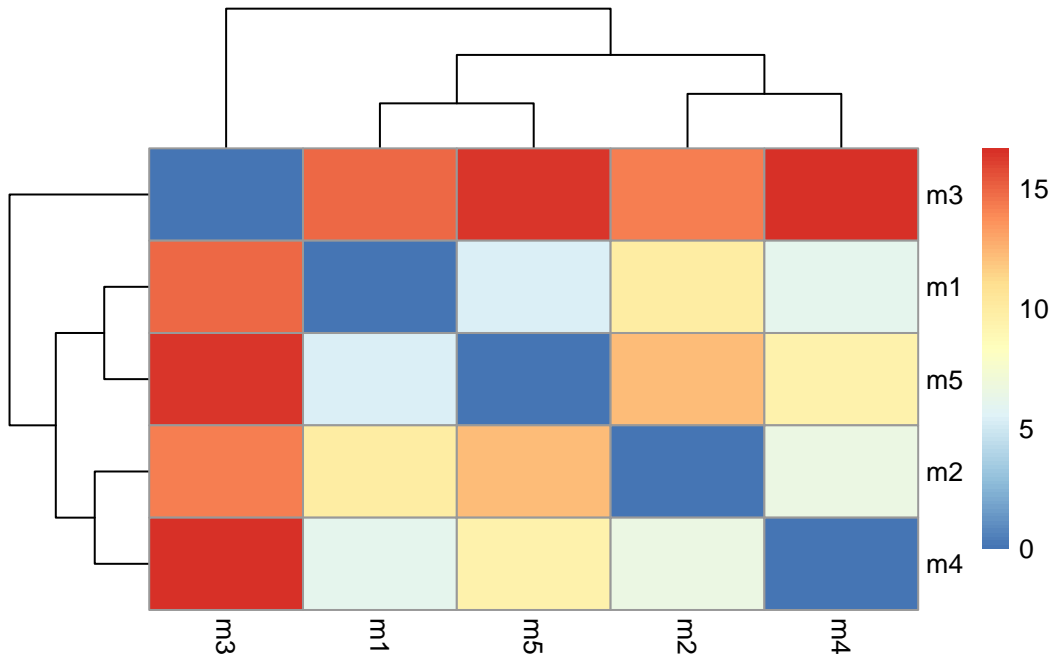
```
range(rd)
```

```
[1] 0.000 16.669
```

```
library(pheatmap)
```

Warning: package 'pheatmap' was built under R version 4.3.3

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



```
# Read a reference PDB structure  
pdb <- read.pdb("1hsg")
```

Note: Accessing on-line PDB file

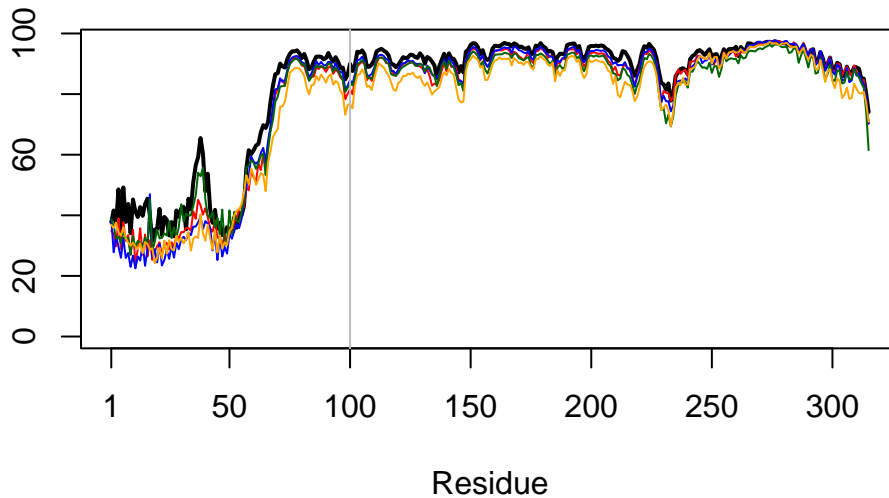
```
plotb3(pdb$b[1,], typ="l", lwd=2, sse=pdb)
```

Warning in plotb3(pdb\$b[1,], typ = "l", lwd = 2, sse = pdb): Length of input 'sse' does not equal the length of input 'x'; Ignoring 'sse'

```

points(pdbb$b[2,], typ="l", col="red")
points(pdbb$b[3,], typ="l", col="blue")
points(pdbb$b[4,], typ="l", col="darkgreen")
points(pdbb$b[5,], typ="l", col="orange")
abline(v=100, col="gray")

```



```

core <- core.find(pdbb)

```

```

core size 314 of 315  vol = 33287.25
core size 313 of 315  vol = 32453.15
core size 312 of 315  vol = 31797.23
core size 311 of 315  vol = 31620.59
core size 310 of 315  vol = 30922.65
core size 309 of 315  vol = 30304.82
core size 308 of 315  vol = 29808.41
core size 307 of 315  vol = 29245.58
core size 306 of 315  vol = 29015.39
core size 305 of 315  vol = 28390.99
core size 304 of 315  vol = 27876.03
core size 303 of 315  vol = 27463.15
core size 302 of 315  vol = 26795.03
core size 301 of 315  vol = 26445.58

```

core size 300 of 315 vol = 25757.37
core size 299 of 315 vol = 25095.67
core size 298 of 315 vol = 24425.66
core size 297 of 315 vol = 23565.5
core size 296 of 315 vol = 22650.02
core size 295 of 315 vol = 21857.43
core size 294 of 315 vol = 20888.56
core size 293 of 315 vol = 20198.7
core size 292 of 315 vol = 19422.06
core size 291 of 315 vol = 18305.19
core size 290 of 315 vol = 17474.97
core size 289 of 315 vol = 16701.25
core size 288 of 315 vol = 15924.72
core size 287 of 315 vol = 15169.24
core size 286 of 315 vol = 14464.79
core size 285 of 315 vol = 13600.42
core size 284 of 315 vol = 12784.32
core size 283 of 315 vol = 11812.13
core size 282 of 315 vol = 10792.69
core size 281 of 315 vol = 9835.128
core size 280 of 315 vol = 8711.522
core size 279 of 315 vol = 7646.416
core size 278 of 315 vol = 6573.881
core size 277 of 315 vol = 5559.989
core size 276 of 315 vol = 4378.986
core size 275 of 315 vol = 3386.086
core size 274 of 315 vol = 2635.794
core size 273 of 315 vol = 1853.574
core size 272 of 315 vol = 1337.132
core size 271 of 315 vol = 963.288
core size 270 of 315 vol = 705.467
core size 269 of 315 vol = 466.324
core size 268 of 315 vol = 282.623
core size 267 of 315 vol = 174.153
core size 266 of 315 vol = 101.963
core size 265 of 315 vol = 58.544
core size 264 of 315 vol = 46.005
core size 263 of 315 vol = 36.91
core size 262 of 315 vol = 29.45
core size 261 of 315 vol = 22.198
core size 260 of 315 vol = 18.608
core size 259 of 315 vol = 16.527
core size 258 of 315 vol = 15.083

core size	257 of 315	vol = 14.169
core size	256 of 315	vol = 13.301
core size	255 of 315	vol = 12.7
core size	254 of 315	vol = 12.271
core size	253 of 315	vol = 11.728
core size	252 of 315	vol = 11.165
core size	251 of 315	vol = 10.633
core size	250 of 315	vol = 10.163
core size	249 of 315	vol = 9.758
core size	248 of 315	vol = 9.358
core size	247 of 315	vol = 9.002
core size	246 of 315	vol = 8.746
core size	245 of 315	vol = 8.427
core size	244 of 315	vol = 8.225
core size	243 of 315	vol = 8.022
core size	242 of 315	vol = 7.75
core size	241 of 315	vol = 7.493
core size	240 of 315	vol = 7.248
core size	239 of 315	vol = 7.02
core size	238 of 315	vol = 6.829
core size	237 of 315	vol = 6.679
core size	236 of 315	vol = 6.559
core size	235 of 315	vol = 6.415
core size	234 of 315	vol = 6.241
core size	233 of 315	vol = 6.063
core size	232 of 315	vol = 5.917
core size	231 of 315	vol = 5.778
core size	230 of 315	vol = 5.634
core size	229 of 315	vol = 5.508
core size	228 of 315	vol = 5.383
core size	227 of 315	vol = 5.218
core size	226 of 315	vol = 5.124
core size	225 of 315	vol = 4.976
core size	224 of 315	vol = 4.879
core size	223 of 315	vol = 4.787
core size	222 of 315	vol = 4.627
core size	221 of 315	vol = 4.487
core size	220 of 315	vol = 4.41
core size	219 of 315	vol = 4.3
core size	218 of 315	vol = 4.201
core size	217 of 315	vol = 4.098
core size	216 of 315	vol = 4.049
core size	215 of 315	vol = 3.964

core size	214 of 315	vol = 3.936
core size	213 of 315	vol = 3.927
core size	212 of 315	vol = 3.908
core size	211 of 315	vol = 3.827
core size	210 of 315	vol = 3.777
core size	209 of 315	vol = 3.722
core size	208 of 315	vol = 3.706
core size	207 of 315	vol = 3.677
core size	206 of 315	vol = 3.672
core size	205 of 315	vol = 3.649
core size	204 of 315	vol = 3.627
core size	203 of 315	vol = 3.592
core size	202 of 315	vol = 3.563
core size	201 of 315	vol = 3.523
core size	200 of 315	vol = 3.49
core size	199 of 315	vol = 3.432
core size	198 of 315	vol = 3.395
core size	197 of 315	vol = 3.349
core size	196 of 315	vol = 3.3
core size	195 of 315	vol = 3.252
core size	194 of 315	vol = 3.194
core size	193 of 315	vol = 3.143
core size	192 of 315	vol = 3.09
core size	191 of 315	vol = 3.028
core size	190 of 315	vol = 2.966
core size	189 of 315	vol = 2.894
core size	188 of 315	vol = 2.799
core size	187 of 315	vol = 2.641
core size	186 of 315	vol = 2.578
core size	185 of 315	vol = 2.497
core size	184 of 315	vol = 2.43
core size	183 of 315	vol = 2.352
core size	182 of 315	vol = 2.275
core size	181 of 315	vol = 2.195
core size	180 of 315	vol = 2.099
core size	179 of 315	vol = 2.006
core size	178 of 315	vol = 1.904
core size	177 of 315	vol = 1.795
core size	176 of 315	vol = 1.69
core size	175 of 315	vol = 1.585
core size	174 of 315	vol = 1.483
core size	173 of 315	vol = 1.37
core size	172 of 315	vol = 1.312

```

core size 171 of 315  vol = 1.221
core size 170 of 315  vol = 1.156
core size 169 of 315  vol = 1.088
core size 168 of 315  vol = 1.018
core size 167 of 315  vol = 0.962
core size 166 of 315  vol = 0.919
core size 165 of 315  vol = 0.863
core size 164 of 315  vol = 0.823
core size 163 of 315  vol = 0.772
core size 162 of 315  vol = 0.726
core size 161 of 315  vol = 0.697
core size 160 of 315  vol = 0.664
core size 159 of 315  vol = 0.643
core size 158 of 315  vol = 0.619
core size 157 of 315  vol = 0.598
core size 156 of 315  vol = 0.577
core size 155 of 315  vol = 0.555
core size 154 of 315  vol = 0.534
core size 153 of 315  vol = 0.513
core size 152 of 315  vol = 0.501
core size 151 of 315  vol = 0.487
FINISHED: Min vol ( 0.5 ) reached

```

```
core.inds <- print(core, vol=0.5)
```

```

# 152 positions (cumulative volume <= 0.5 Angstrom^3)
  start end length
1    69  82     14
2    85  85      1
3    89  97      9
4    99 100      2
5   102 183     82
6   185 228     44

```

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

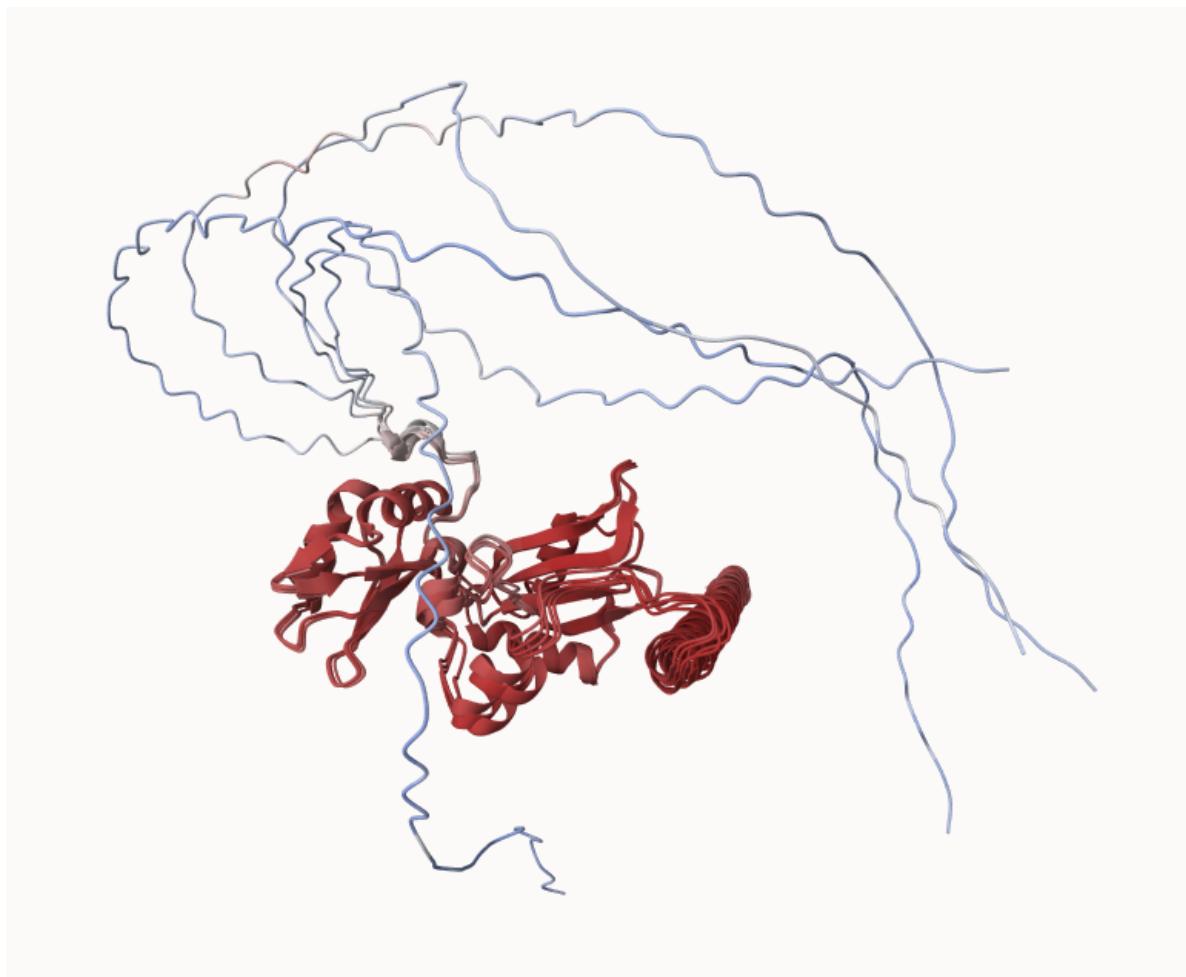
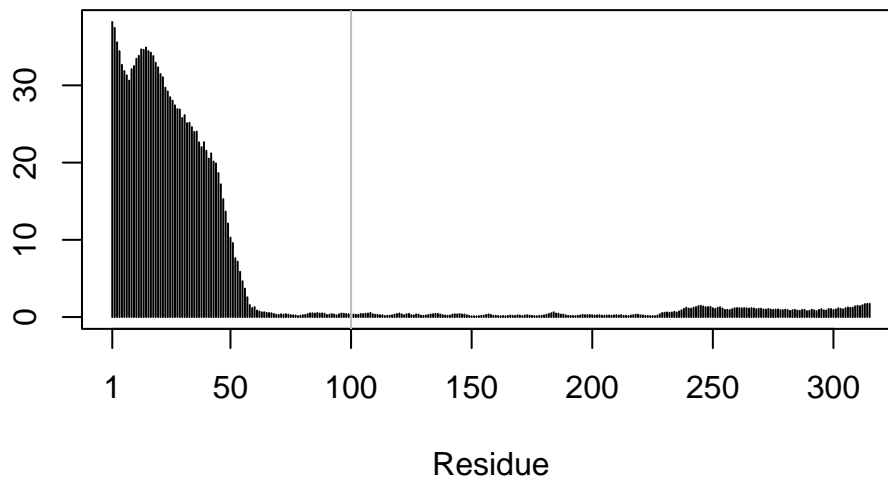


Figure 1: Fig 1: Mol* 3D superimposed image with uncertainty/disordered coloring

```
rf <- rmsf(xyz)  
plotb3(rf, sse=pdb)
```

Warning in plotb3(rf, sse = pdb): Length of input 'sse' does not equal the length of input 'x'; Ignoring 'sse'

```
abline(v=100, col="gray", ylab="RMSF")
```



```
library(jsonlite)
```

Warning: package 'jsonlite' was built under R version 4.3.3

```
# Listing of all PAE JSON files
pae_files <- list.files(path=results_dir,
                        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"
```

```
# Per-residue pLDDT scores
# same as B-factor of PDB..
head(pae1$plddt)
```

```
[1] 37.75 41.59 37.62 48.56 32.00 49.19
```

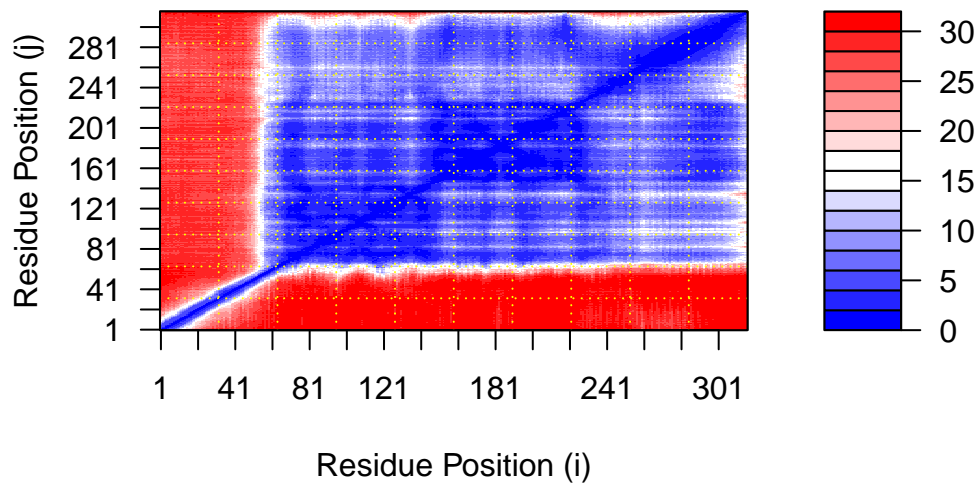
```
pae1$max_pae
```

```
[1] 31.10938
```

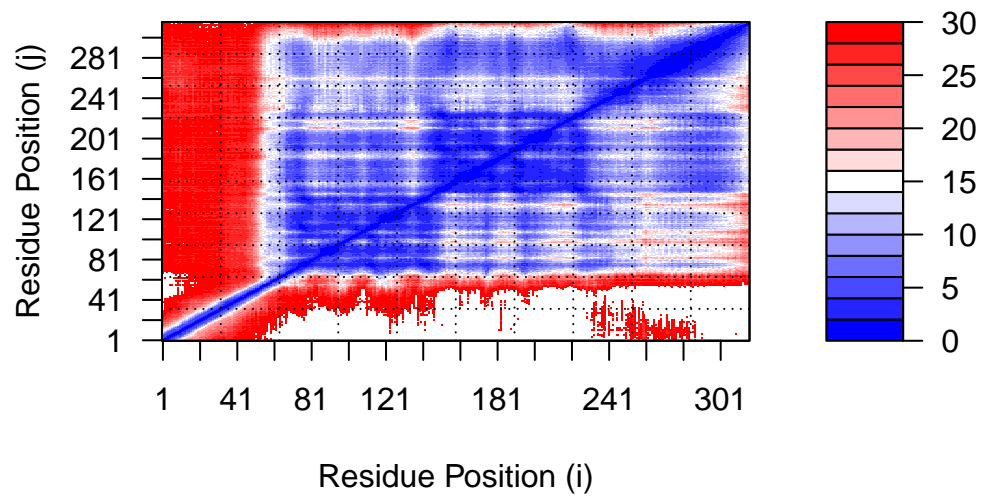
```
pae5$max_pae
```

```
[1] 31.51562
```

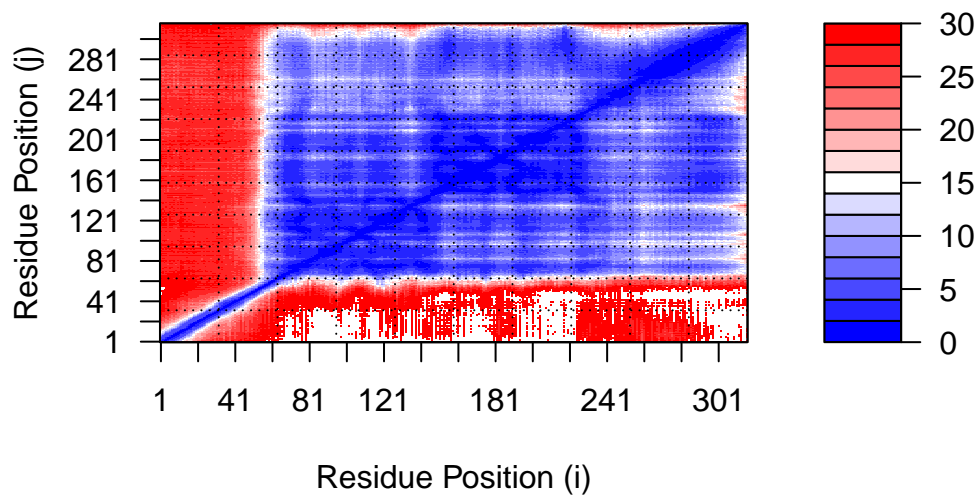
```
plot.dmat(pae1$pae,
          xlab="Residue Position (i)",
          ylab="Residue Position (j)")
```



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



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



Residue conservation from alignment file

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

```
[1] "gene_project_cDNA_f939f/gene_project_cDNA_f939f.a3m"
```

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

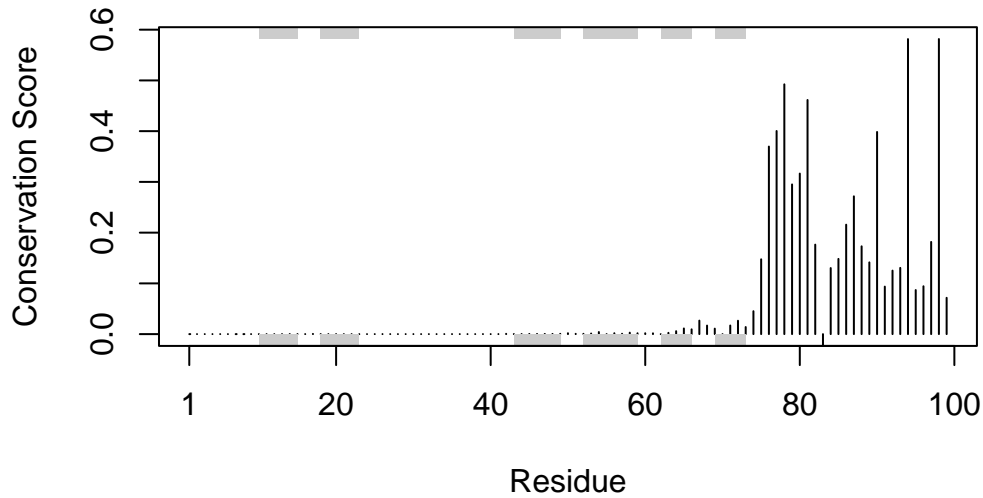
```
[1] " ** Duplicated sequence id's: 101 **"
```

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

```
[1] 13819  461
```

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

```
plotb3(sim[1:99], sse=trim.pdb(pdb, chain="A"),
       ylab="Conservation Score")
```



```
con <- consensus(aln, cutoff = 0.7)
con$seq
```

```
[1] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[19] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[37] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[55] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[73] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[91] "-" "-" "-" "F" "-" "-" "-" "G" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[109] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[127] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[145] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[163] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[181] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[199] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[217] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[235] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[253] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[271] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
```


[289] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[307] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[325] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[343] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[361] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[379] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[397] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[415] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[433] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[451] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"