

AlphaFold assignment

David Alvarez

```
AlphaFold <- "hivprimer_results"
```

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

```
basename(pdb_files)
```

```
[1] "hivprdimer_94b5b_0_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000.pdb"
[2] "hivprdimer_94b5b_0_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000.pdb"
[3] "hivprdimer_94b5b_0_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000.pdb"
[4] "hivprdimer_94b5b_0_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000.pdb"
[5] "hivprdimer_94b5b_0_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000.pdb"
```

```
library(bio3d)
```

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

Reading PDB files:

```
hivprimer_results/hivprdimer_94b5b_0_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000.pdb
hivprimer_results/hivprdimer_94b5b_0_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000.pdb
hivprimer_results/hivprdimer_94b5b_0_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000.pdb
hivprimer_results/hivprdimer_94b5b_0_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000.pdb
hivprimer_results/hivprdimer_94b5b_0_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000.pdb
.....
```

Extracting sequences

```
pdb/seq: 1    name: hivprimer_results/hivprdimer_94b5b_0_unrelaxed_rank_001_alphafold2_ptm_mo
pdb/seq: 2    name: hivprimer_results/hivprdimer_94b5b_0_unrelaxed_rank_002_alphafold2_ptm_mo
pdb/seq: 3    name: hivprimer_results/hivprdimer_94b5b_0_unrelaxed_rank_003_alphafold2_ptm_mo
```

```

pdb/seq: 4    name: hivprimer_results/hivprdimer_94b5b_0_unrelaxed_rank_004_alphafold2_ptm_mo
pdb/seq: 5    name: hivprimer_results/hivprdimer_94b5b_0_unrelaxed_rank_005_alphafold2_ptm_mo

```

```
pdbs
```

```

1          .          .          .          .          50
[Truncated_Name:1]hivprdimer PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGI
[Truncated_Name:2]hivprdimer PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGI
[Truncated_Name:3]hivprdimer PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGI
[Truncated_Name:4]hivprdimer PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGI
[Truncated_Name:5]hivprdimer PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGI
*****
1          .          .          .          .          50

51          .          .          .          .          99
[Truncated_Name:1]hivprdimer GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:2]hivprdimer GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:3]hivprdimer GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:4]hivprdimer GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:5]hivprdimer GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
*****
51          .          .          .          .          99

```

Call:

```
pdbaln(files = pdb_files, fit = TRUE, exefile = "msa")
```

Class:

```
pdbs, fasta
```

Alignment dimensions:

```
5 sequence rows; 99 position columns (99 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 99 non NA positions

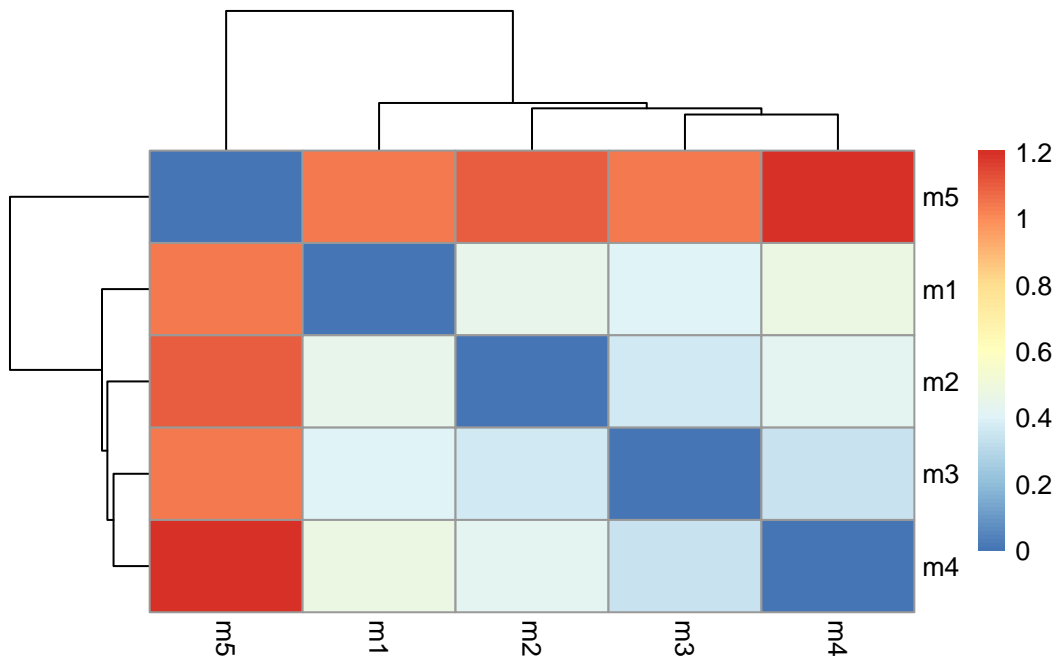
```
range(rd)
```

```
[1] 0.000 1.205
```

```
library(pheatmap)
```

Warning: package 'pheatmap' was built under R version 4.3.2

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



```
pdb <- read.pdb("1hsg")
```

Note: Accessing on-line PDB file

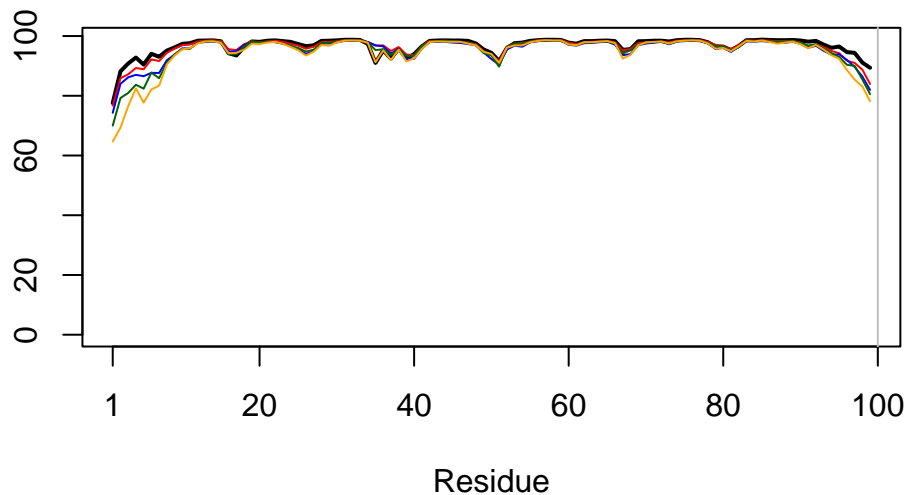
```
plotb3(pdb$b[1,], typ="l", lwd=2, sse=pdb)
```

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 98 of 99  vol = 3.444
core size 97 of 99  vol = 2.62
core size 96 of 99  vol = 2.159
core size 95 of 99  vol = 1.66
core size 94 of 99  vol = 1.312
core size 93 of 99  vol = 1.014
core size 92 of 99  vol = 0.762
core size 91 of 99  vol = 0.566
core size 90 of 99  vol = 0.409
FINISHED: Min vol ( 0.5 ) reached

```

```

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

```

```
# 91 positions (cumulative volume <= 0.5 Angstrom^3)
  start end length
1     3   3     1
2     7  96    90
```

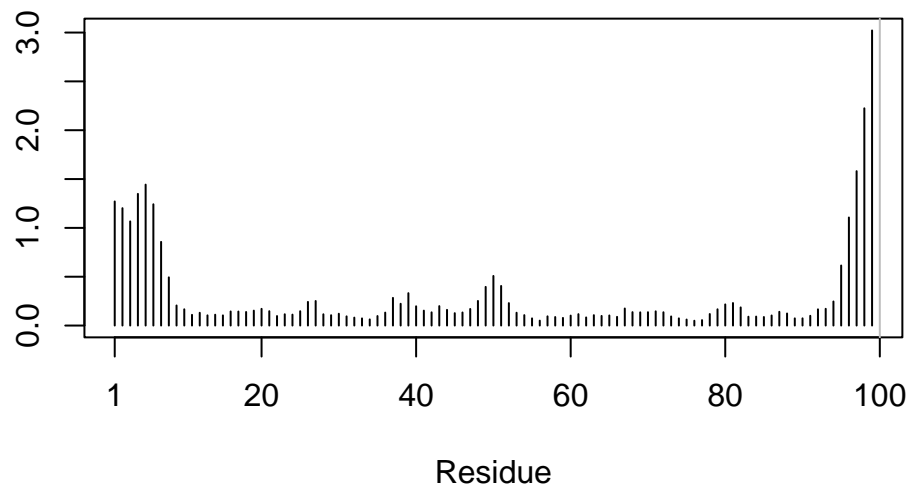
```
xyz <- pdbfit(pdb, core.inds, outpath="corefit_structures")
```

```
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 4.3.2

```
pae_files <- list.files(path="json files", 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"
```

```
head(pae1$plddt)
```

```
[1] 77.56 88.19 90.81 92.81 90.44 94.06
```

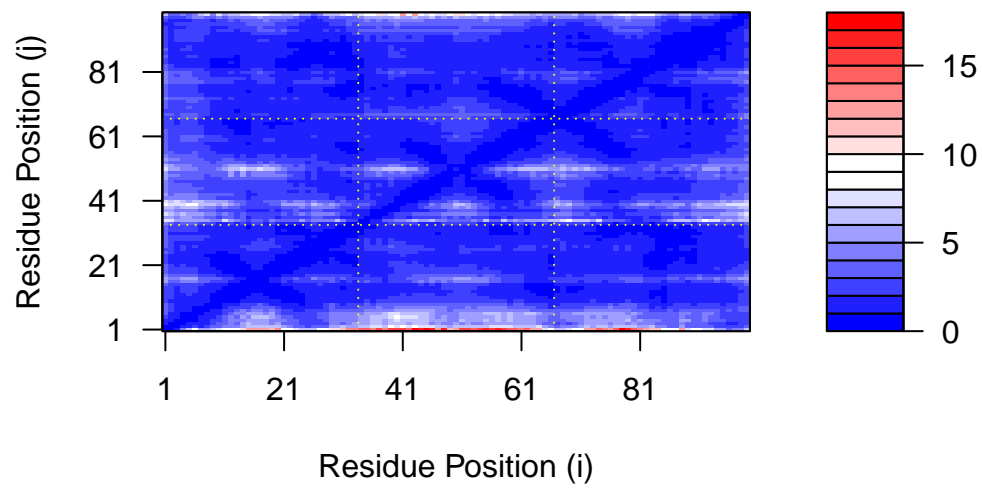
```
pae1$max_pae
```

```
[1] 17.65625
```

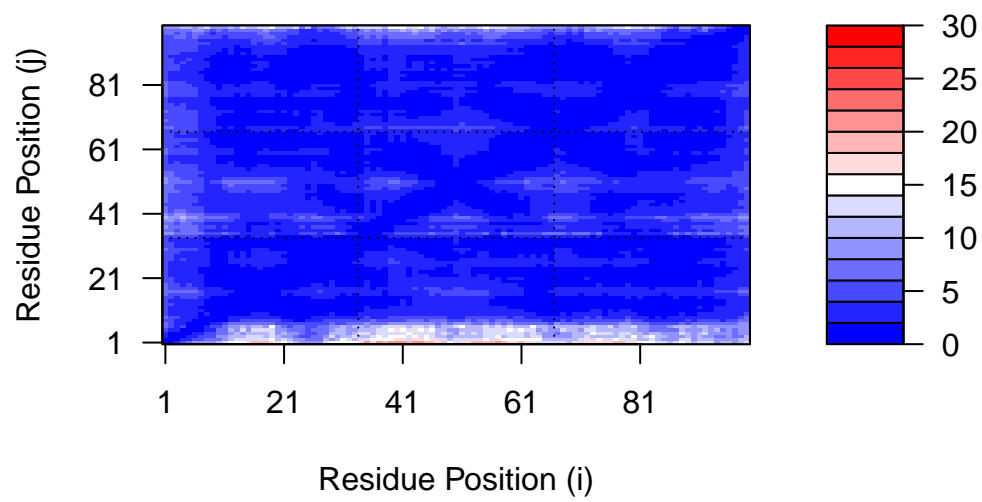
```
pae5$max_pae
```

```
[1] 21.1875
```

```
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))
```