Class 11

Jacqueline Cheung(A17085191)

```
# Change this for YOUR results dir name
results_dir <- "hivprdimer_23119/"
```

```
# File names for all PDB models
pdb_files <- list.files(path=results_dir,
                        pattern="*.pdb",
                        full.names = TRUE)

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

```
[1] "hivprdimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000.pdb"
[2] "hivprdimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000.pdb"
[3] "hivprdimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000.pdb"
[4] "hivprdimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_000.pdb"
[5] "hivprdimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000.pdb"
```

```
library(bio3d)

# Read all data from Models
#  and superpose/fit coords
pdbs <- pdbaln(pdb_files, fit=TRUE, exefile="msa")
```

```
Reading PDB files:
hivprdimer_23119/hivprdimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000
hivprdimer_23119/hivprdimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
hivprdimer_23119/hivprdimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
hivprdimer_23119/hivprdimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_000
hivprdimer_23119/hivprdimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000
.....
```

```
Extracting sequences

pdb/seq: 1   name: hivprdimer_23119/hivprdimer_23119_unrelaxed_rank_001_alphafold2_multimer_v
pdb/seq: 2   name: hivprdimer_23119/hivprdimer_23119_unrelaxed_rank_002_alphafold2_multimer_v
pdb/seq: 3   name: hivprdimer_23119/hivprdimer_23119_unrelaxed_rank_003_alphafold2_multimer_v
pdb/seq: 4   name: hivprdimer_23119/hivprdimer_23119_unrelaxed_rank_004_alphafold2_multimer_v
pdb/seq: 5   name: hivprdimer_23119/hivprdimer_23119_unrelaxed_rank_005_alphafold2_multimer_v
```

```
pdbs
```

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

                             51        .         .         .         .         100
[Truncated_Name:1]hivprdimer   GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:2]hivprdimer   GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:3]hivprdimer   GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:4]hivprdimer   GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:5]hivprdimer   GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
                               **************************************************
                             51        .         .         .         .         100

                            101        .         .         .         .         150
[Truncated_Name:1]hivprdimer   QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIG
[Truncated_Name:2]hivprdimer   QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIG
[Truncated_Name:3]hivprdimer   QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIG
[Truncated_Name:4]hivprdimer   QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIG
[Truncated_Name:5]hivprdimer   QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIG
                               **************************************************
                            101        .         .         .         .         150

                            151        .         .         .         .       198
[Truncated_Name:1]hivprdimer   GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:2]hivprdimer   GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:3]hivprdimer   GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
```

```
[Truncated_Name:4]hivprdimer   GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:5]hivprdimer   GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
                               ********************************************
                             151        .         .         .         .       198

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

Class:
  pdbs, fasta

Alignment dimensions:
  5 sequence rows; 198 position columns (198 non-gap, 0 gap)

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

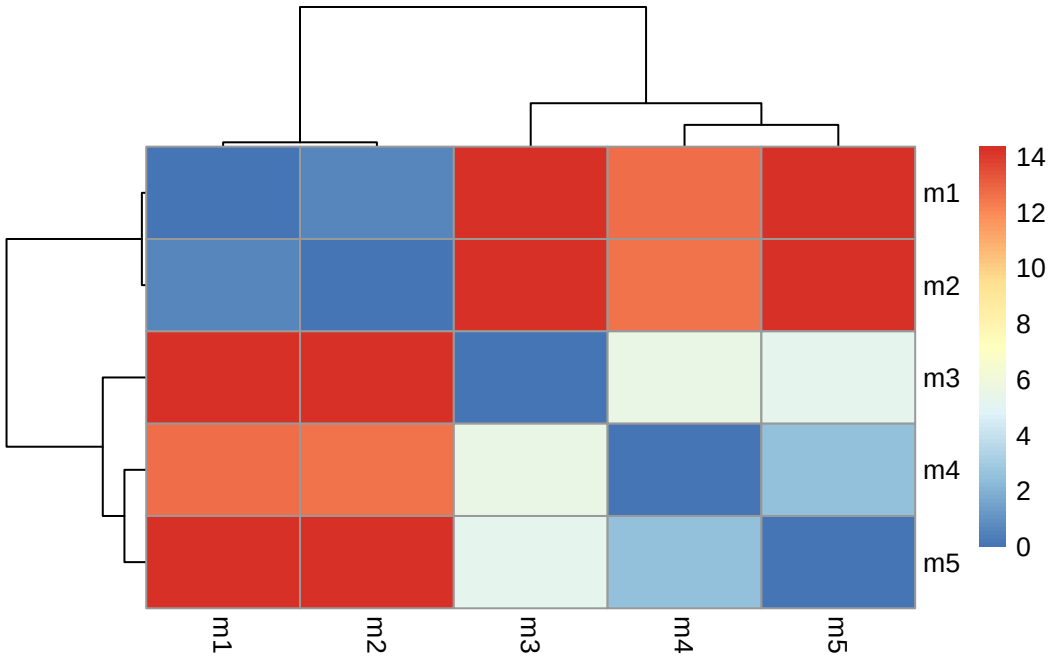
```
rd <- rmsd(pdbs, fit=T)
```

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

```
range(rd)
```

```
[1]  0.000 14.342
```

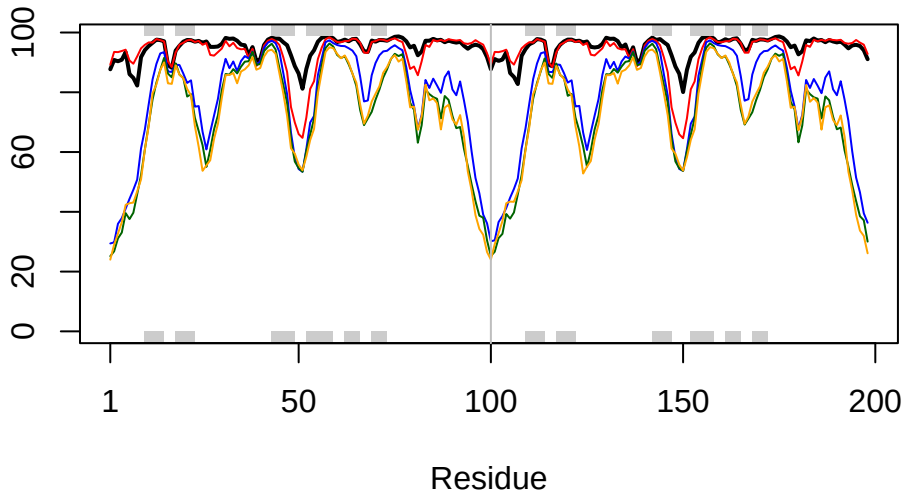
```
library(pheatmap)

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(pdbs$b[1,], typ="l", lwd=2, sse=pdb)
points(pdbs$b[2,], typ="l", col="red")
points(pdbs$b[3,], typ="l", col="blue")
points(pdbs$b[4,], typ="l", col="darkgreen")
points(pdbs$b[5,], typ="l", col="orange")
abline(v=100, col="gray")
```




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

```
 core size 197 of 198  vol = 4969.063
 core size 196 of 198  vol = 4361.131
 core size 195 of 198  vol = 4144.622
 core size 194 of 198  vol = 3945.204
 core size 193 of 198  vol = 3747.649
 core size 192 of 198  vol = 3579.207
 core size 191 of 198  vol = 3464.153
 core size 190 of 198  vol = 3336.583
 core size 189 of 198  vol = 3230.417
 core size 188 of 198  vol = 3128.605
 core size 187 of 198  vol = 3067.873
 core size 186 of 198  vol = 3020.412
 core size 185 of 198  vol = 2958.324
 core size 184 of 198  vol = 2909.025
 core size 183 of 198  vol = 2872.84
 core size 182 of 198  vol = 2837.745
 core size 181 of 198  vol = 2845.792
 core size 180 of 198  vol = 2862.112
 core size 179 of 198  vol = 2899.84
 core size 178 of 198  vol = 2951.96
```

```
core size 177 of 198  vol = 3021.471
core size 176 of 198  vol = 3089.976
core size 175 of 198  vol = 3113.405
core size 174 of 198  vol = 3132.11
core size 173 of 198  vol = 3137.044
core size 172 of 198  vol = 3094.072
core size 171 of 198  vol = 3041.709
core size 170 of 198  vol = 2952.307
core size 169 of 198  vol = 2889.611
core size 168 of 198  vol = 2831.304
core size 167 of 198  vol = 2748.293
core size 166 of 198  vol = 2672.553
core size 165 of 198  vol = 2601.517
core size 164 of 198  vol = 2534.316
core size 163 of 198  vol = 2463.38
core size 162 of 198  vol = 2388.514
core size 161 of 198  vol = 2320.646
core size 160 of 198  vol = 2235.029
core size 159 of 198  vol = 2160.253
core size 158 of 198  vol = 2077.504
core size 157 of 198  vol = 2003.263
core size 156 of 198  vol = 1938.643
core size 155 of 198  vol = 1857.372
core size 154 of 198  vol = 1780.914
core size 153 of 198  vol = 1700.584
core size 152 of 198  vol = 1626.794
core size 151 of 198  vol = 1550.207
core size 150 of 198  vol = 1477.57
core size 149 of 198  vol = 1419.981
core size 148 of 198  vol = 1360.005
core size 147 of 198  vol = 1302.722
core size 146 of 198  vol = 1253.857
core size 145 of 198  vol = 1211.167
core size 144 of 198  vol = 1168.585
core size 143 of 198  vol = 1118.03
core size 142 of 198  vol = 1071.861
core size 141 of 198  vol = 1035.74
core size 140 of 198  vol = 993.591
core size 139 of 198  vol = 952.159
core size 138 of 198  vol = 906.029
core size 137 of 198  vol = 863.817
core size 136 of 198  vol = 838.135
core size 135 of 198  vol = 806.348
```

```
core size 134 of 198  vol = 767.994
core size 133 of 198  vol = 730.361
core size 132 of 198  vol = 694.175
core size 131 of 198  vol = 652.458
core size 130 of 198  vol = 615.817
core size 129 of 198  vol = 578.416
core size 128 of 198  vol = 544.624
core size 127 of 198  vol = 516.657
core size 126 of 198  vol = 489.752
core size 125 of 198  vol = 454.196
core size 124 of 198  vol = 430.905
core size 123 of 198  vol = 416.433
core size 122 of 198  vol = 406.978
core size 121 of 198  vol = 377.755
core size 120 of 198  vol = 365.748
core size 119 of 198  vol = 340.859
core size 118 of 198  vol = 307.863
core size 117 of 198  vol = 285.017
core size 116 of 198  vol = 265.162
core size 115 of 198  vol = 243.16
core size 114 of 198  vol = 223.911
core size 113 of 198  vol = 200.144
core size 112 of 198  vol = 179.195
core size 111 of 198  vol = 164.095
core size 110 of 198  vol = 149.837
core size 109 of 198  vol = 138.771
core size 108 of 198  vol = 122.917
core size 107 of 198  vol = 108.842
core size 106 of 198  vol = 98.074
core size 105 of 198  vol = 88.841
core size 104 of 198  vol = 80.268
core size 103 of 198  vol = 73.282
core size 102 of 198  vol = 66.192
core size 101 of 198  vol = 61.249
core size 100 of 198  vol = 57.566
core size 99 of 198  vol = 54.35
core size 98 of 198  vol = 49.979
core size 97 of 198  vol = 44.928
core size 96 of 198  vol = 40.613
core size 95 of 198  vol = 33.814
core size 94 of 198  vol = 27.945
core size 93 of 198  vol = 22.83
core size 92 of 198  vol = 16.741
```

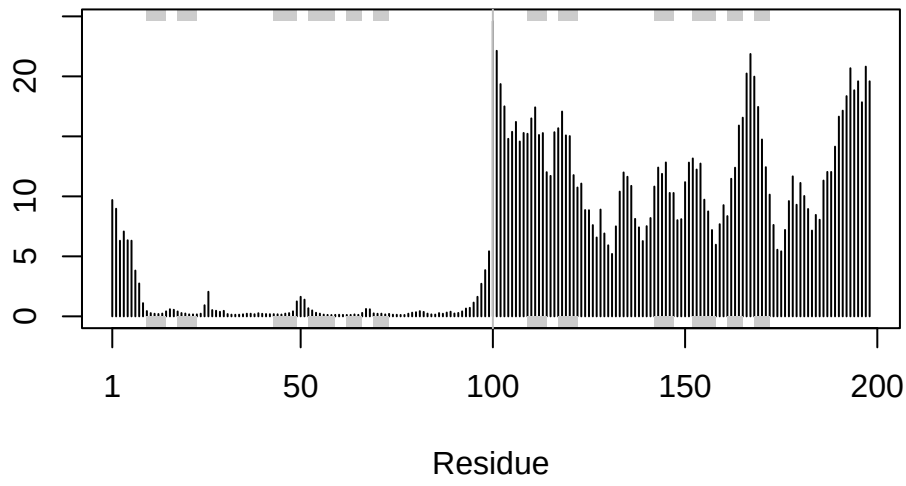
```
 core size 91 of 198  vol = 10.654
 core size 90 of 198  vol = 4.622
 core size 89 of 198  vol = 3.099
 core size 88 of 198  vol = 2.631
 core size 87 of 198  vol = 2.244
 core size 86 of 198  vol = 1.89
 core size 85 of 198  vol = 1.545
 core size 84 of 198  vol = 1.278
 core size 83 of 198  vol = 1.071
 core size 82 of 198  vol = 0.889
 core size 81 of 198  vol = 0.751
 core size 80 of 198  vol = 0.636
 core size 79 of 198  vol = 0.585
 core size 78 of 198  vol = 0.521
 core size 77 of 198  vol = 0.479
 FINISHED: Min vol ( 0.5 ) reached
```

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

```
# 78 positions (cumulative volume <= 0.5 Angstrom^3)
  start end length
1    10  25     16
2    28  48     21
3    53  93     41
```

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

```
rf <- rmsf(xyz)

plotb3(rf, sse=pdb)
abline(v=100, col="gray", ylab="RMSF")
```



Predicted Alignment Error for domains

```
library(jsonlite)

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

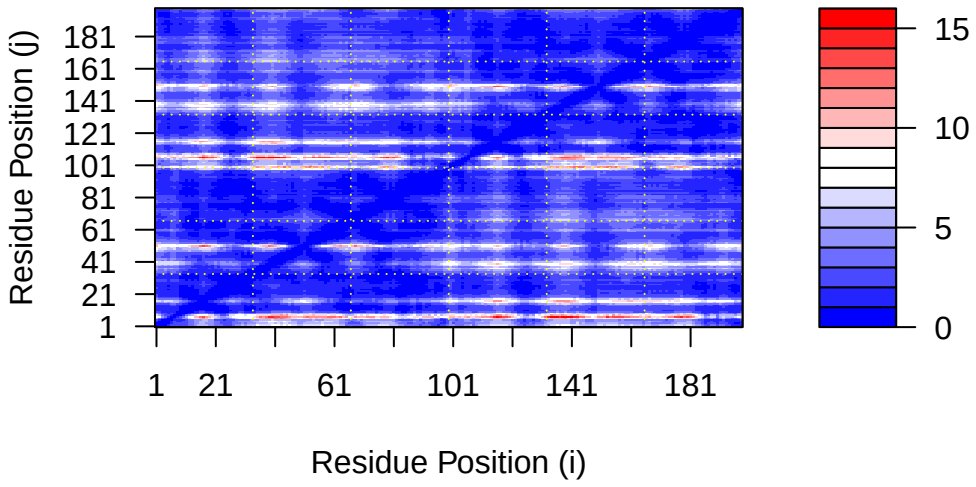
```
pae1 <- read_json(pae_files[1],simplifyVector = TRUE)
pae5 <- read_json(pae_files[5],simplifyVector = TRUE)

attributes(pae1)
```

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

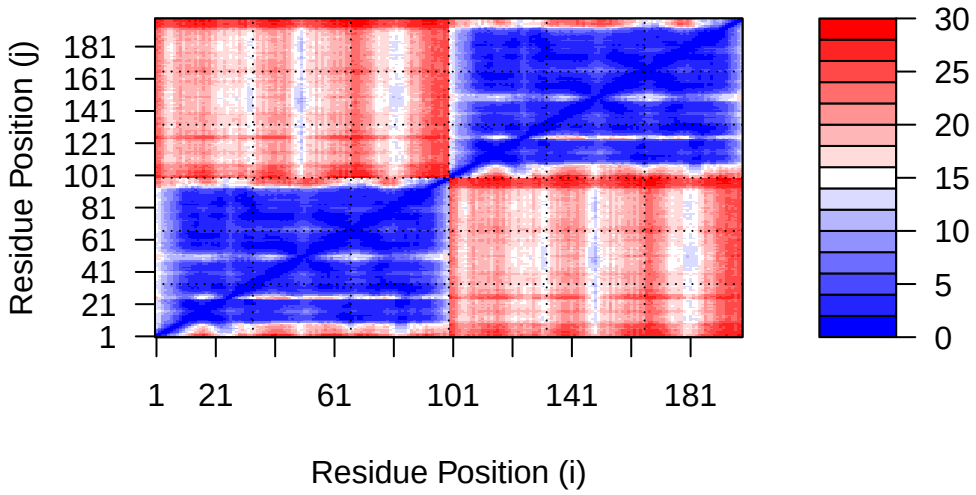
```
# Per-residue pLDDT scores
#  same as B-factor of PDB..
head(pae1$plddt)
```

```
[1] 87.69 90.81 90.38 90.88 93.44 86.06
```

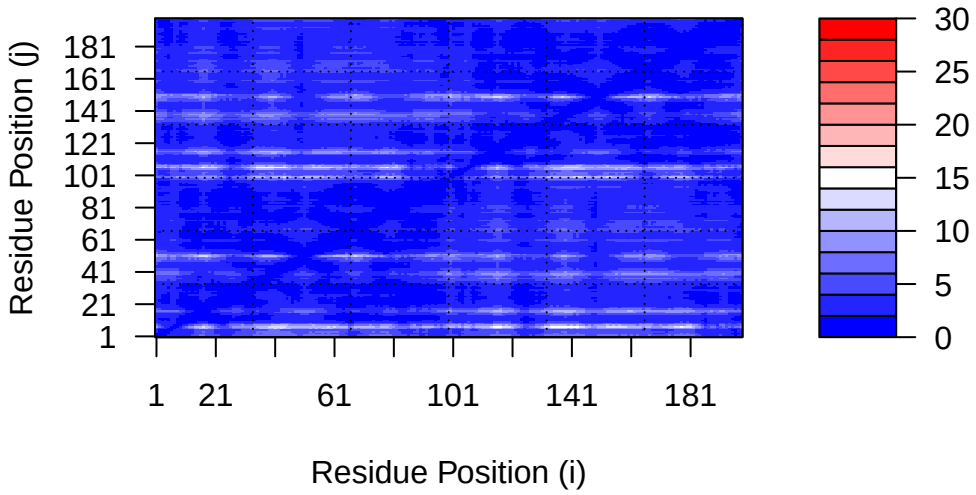
```
plot.dmat(pae1$pae,
          xlab="Residue Position (i)",
          ylab="Residue Position (j)")
```




```
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] "hivprdimer_23119/hivprdimer_23119.a3m"
```

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

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

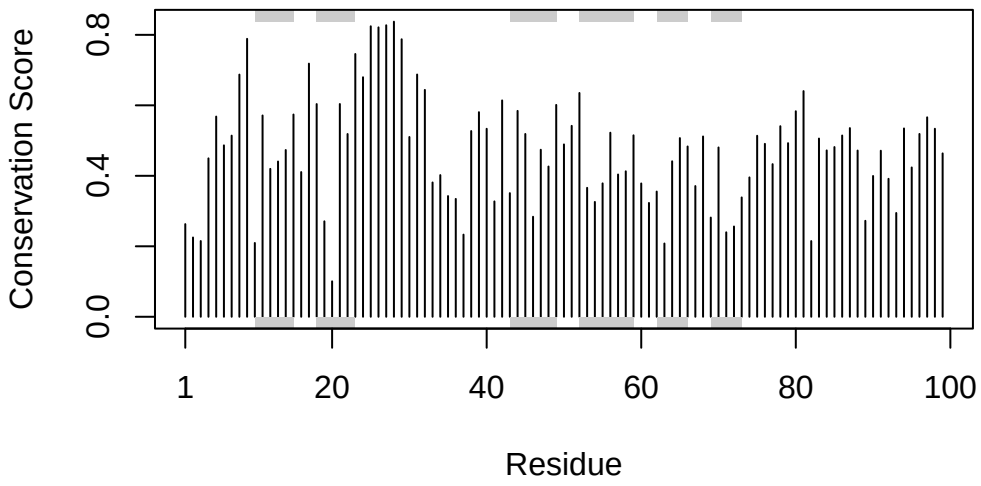
Q. How many sequences are in this alignment?

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

```
[1] 5378  132
```

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

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




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

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

```
m1.pdb <- read.pdb(pdb_files[1])
occ <- vec2resno(c(sim[1:99], sim[1:99]), m1.pdb$atom$resno)
write.pdb(m1.pdb, o=occ, file="m1_conserv.pdb")
```