Structural Bioinformatics (pt.2)

Angela Abraham

2025-11-09

Custom analysis of resulting models

```
dir.exists("Pep.94b5b.result.zip")
```

```
## [1] FALSE
```

```
unzip("Pep_94b5b.result.zip")
list.dirs()
```

```
##  [1] "."
##  [2] "./.Rproj.user"
##  [3] "./.Rproj.user/C60B3194"
##  [4] "./.Rproj.user/C60B3194/bibliography-index"
##  [5] "./.Rproj.user/C60B3194/ctx"
##  [6] "./.Rproj.user/C60B3194/explorer-cache"
##  [7] "./.Rproj.user/C60B3194/pcs"
##  [8] "./.Rproj.user/C60B3194/presentation"
##  [9] "./.Rproj.user/C60B3194/profiles-cache"
## [10] "./.Rproj.user/C60B3194/sources"
## [11] "./.Rproj.user/C60B3194/sources/prop"
## [12] "./.Rproj.user/C60B3194/sources/session-8dfc65f8"
## [13] "./.Rproj.user/C60B3194/tutorial"
## [14] "./.Rproj.user/C60B3194/unsaved-notebooks"
## [15] "./.Rproj.user/C60B3194/unsaved-notebooks/C3660953"
## [16] "./.Rproj.user/C60B3194/viewer-cache"
## [17] "./.Rproj.user/shared"
## [18] "./.Rproj.user/shared/notebooks"
## [19] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report"
## [20] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1"
## [21] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/C60B31948dfc65f8"
## [22] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s"
## [23] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/c01jyn9fsdt4f"
## [24] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/c0jdp9o49ot6u"
## [25] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/c182x7rz3914d"
## [26] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/c2pp2u4ufbm4y"
## [27] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/c4nmcjwrrzfaw"
## [28] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/c8ngje5yf3dvo"
## [29] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/c8q6fugvu29sm"
```

```
## [30] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/c9cd7mfkj8iyw"
## [31] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/ca9qdn6phwyeo"
## [32] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/cau8293ue8nht"
## [33] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/cau8293ue8nht/temp"
## [34] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/cb35mdn75xgu8"
## [35] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/cbw3qatrf5cie"
## [36] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/cc5y1rhrgt1g5"
## [37] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/ccfgndsulhy2u"
## [38] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/ccjqyphhqqqfb"
## [39] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/ceoyvq68hsqmk"
## [40] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/ch0x11ybtjhti"
## [41] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/cht8h9n6waoli"
## [42] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/chutd6wcrgcqd"
## [43] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/cjvrzr8ilvv8h"
## [44] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/cmb9e2uhkr0mv"
## [45] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/cnksaj16uye5b"
## [46] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/cocqtoy28txis"
## [47] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/cqo5y6t51p0gn"
## [48] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/cte8v49l0n01h"
## [49] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/ctoga21v5opoo"
## [50] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/cuh6g7tcbb2g6"
## [51] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/cyd1qk95dpxa9"
## [52] "./.Rproj.user/shared/notebooks/F4C9C562-11_Lab_Report/1/s/czjllc5kzh4kj"
## [53] "./11_Lab_Report_files"
## [54] "./11_Lab_Report_files/figure-latex"
## [55] "./corefit_structures"
## [56] "./Pep_94b5b"
## [57] "./Pep_94b5b/Pep_94b5b_env"
```

```
list.dirs(recursive=FALSE)
```

```
## [1] "./.Rproj.user"         "./11_Lab_Report_files" "./corefit_structures"
## [4] "./Pep_94b5b"
```

```
results_dir<- "./Pep_94b5b"
library(bio3d)
```

```
## Warning: package 'bio3d' was built under R version 4.5.2
```

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

```
print(paste("Found", length(pdb_files), "PDB files"))
```

```
## [1] "Found 5 PDB files"
```

```
basename(pdb_files)
```

```
## [1] "Pep_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_4_seed_000.pdb"
## [2] "Pep_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_5_seed_000.pdb"
## [3] "Pep_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_3_seed_000.pdb"
## [4] "Pep_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_1_seed_000.pdb"
## [5] "Pep_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000.pdb"
```

```
library(bio3d)
pdbs<-pdbaln(pdb_files, fit=TRUE, exefile="msa")
```

```
## Reading PDB files:
## ./Pep_94b5b/Pep_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_4_seed_000.pdb
## ./Pep_94b5b/Pep_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_5_seed_000.pdb
## ./Pep_94b5b/Pep_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_3_seed_000.pdb
## ./Pep_94b5b/Pep_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_1_seed_000.pdb
## ./Pep_94b5b/Pep_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000.pdb
## .....
##
## Extracting sequences
##
## pdb/seq: 1   name: ./Pep_94b5b/Pep_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_4_seed_000.pdb
## pdb/seq: 2   name: ./Pep_94b5b/Pep_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_5_seed_000.pdb
## pdb/seq: 3   name: ./Pep_94b5b/Pep_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_3_seed_000.pdb
## pdb/seq: 4   name: ./Pep_94b5b/Pep_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_1_seed_000.pdb
## pdb/seq: 5   name: ./Pep_94b5b/Pep_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000.pdb
```

```
pdbs
```

```
##                                1        .         .         .         .         50
## [Truncated_Name:1]Pep_94b5b_   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
## [Truncated_Name:2]Pep_94b5b_   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
## [Truncated_Name:3]Pep_94b5b_   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
## [Truncated_Name:4]Pep_94b5b_   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
## [Truncated_Name:5]Pep_94b5b_   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
##                                **************************************************
##                                1        .         .         .         .         50
##
##                               51        .         .         .         .        99
## [Truncated_Name:1]Pep_94b5b_   GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
## [Truncated_Name:2]Pep_94b5b_   GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
## [Truncated_Name:3]Pep_94b5b_   GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
## [Truncated_Name:4]Pep_94b5b_   GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
## [Truncated_Name:5]Pep_94b5b_   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(pdbs, fit=T)
```

```
## Warning in rmsd(pdbs, fit = T): No indices provided, using the 99 non NA positions
```

```
range(rd)
```

```
## [1] 0.00 1.37
```

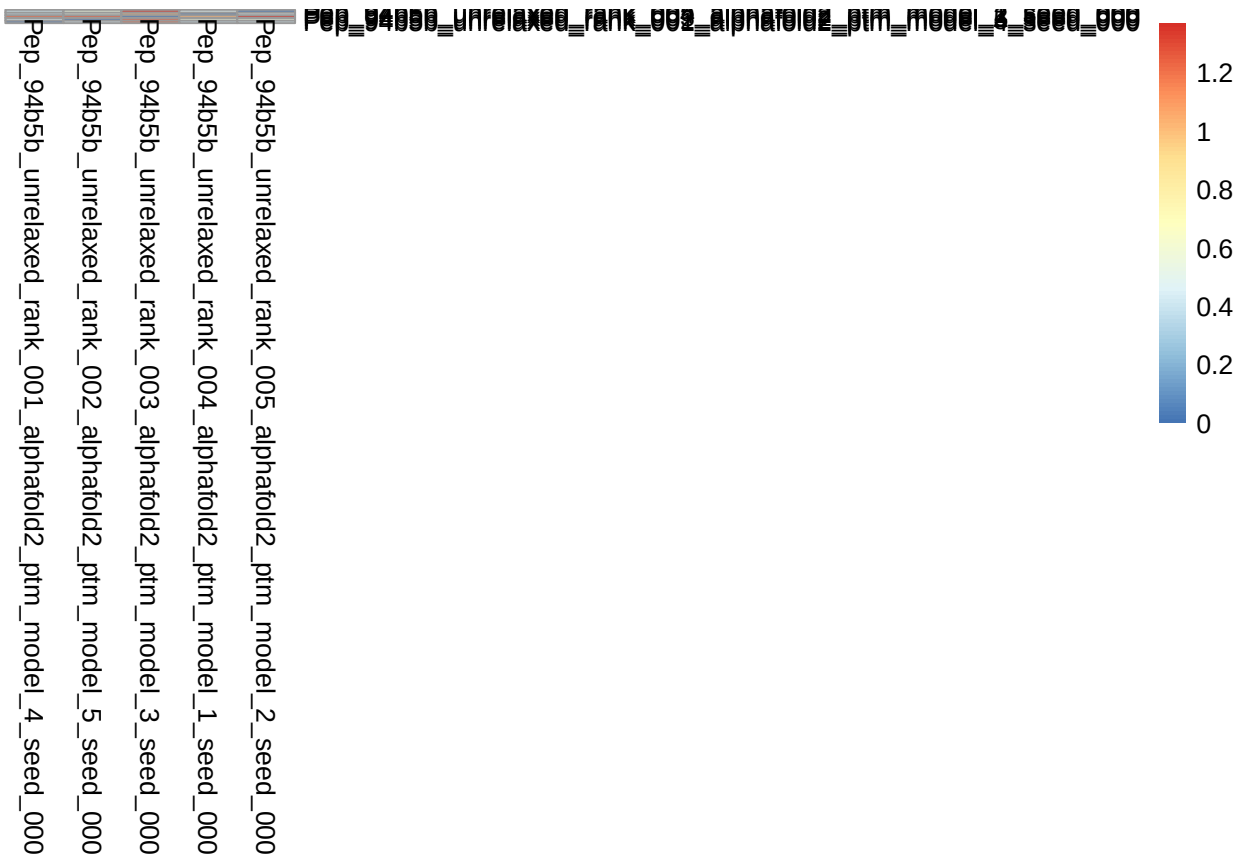
```
library(bio3d)
library(pheatmap)
```

```
## Warning: package 'pheatmap' was built under R version 4.5.2
```

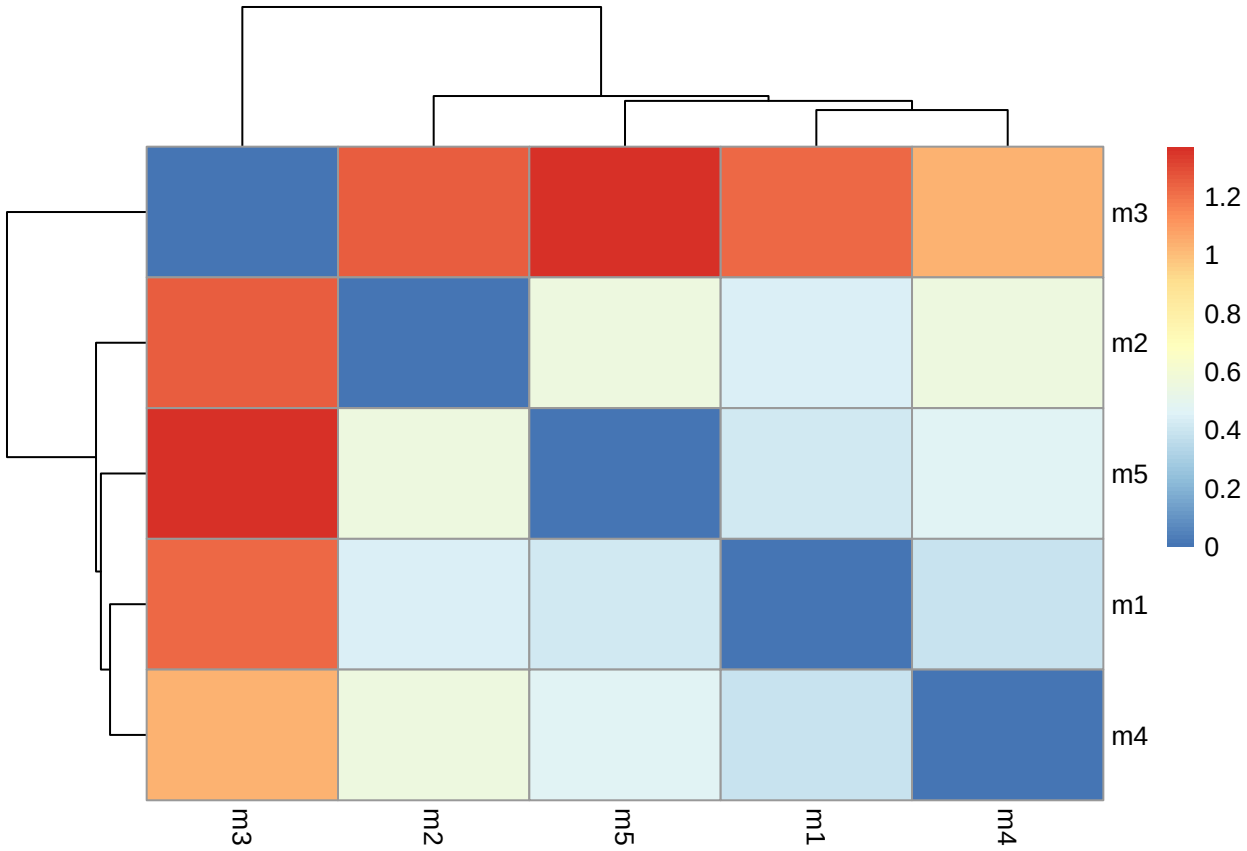
```
rdist<-rmsd(pdbs)
```

```
## Warning in rmsd(pdbs): No indices provided, using the 99 non NA positions
```

```
pheatmap(rdist,
      cluster_rows=FALSE,
      cluster_cols=FALSE)
```




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



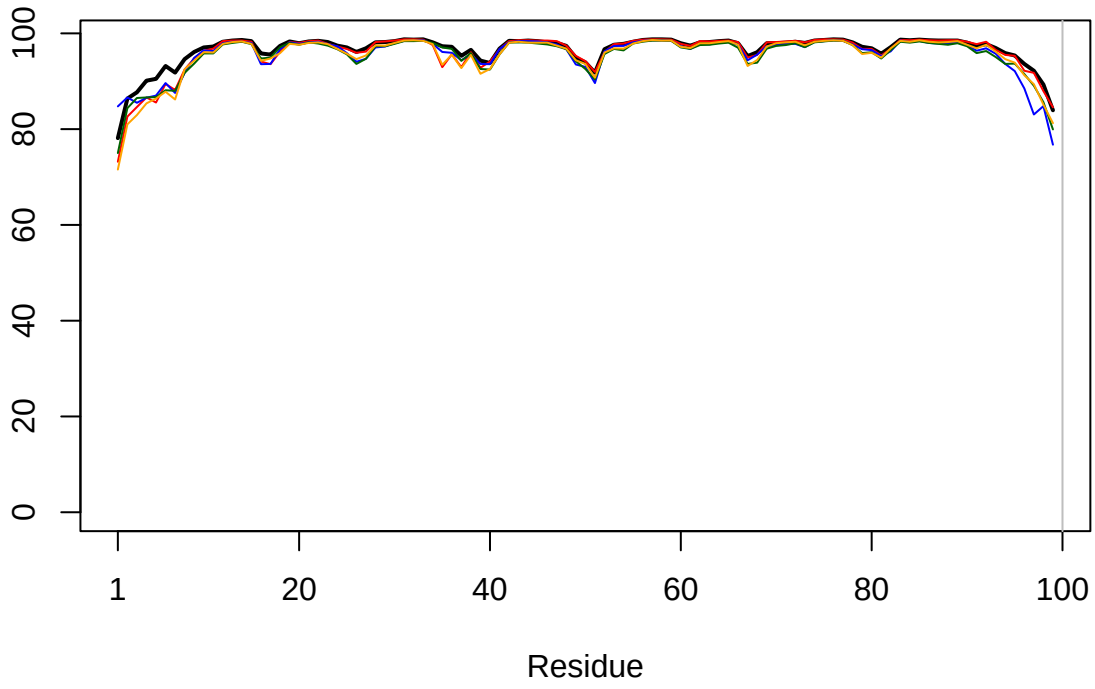

```
pdb<-read.pdb("1hsg")
```

```
##   Note: Accessing on-line PDB file
```

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

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

```
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 98 of 99  vol = 4.376
##  core size 97 of 99  vol = 3.604
##  core size 96 of 99  vol = 2.86
##  core size 95 of 99  vol = 2.352
##  core size 94 of 99  vol = 1.794
##  core size 93 of 99  vol = 1.303
##  core size 92 of 99  vol = 0.946
##  core size 91 of 99  vol = 0.579
##  core size 90 of 99  vol = 0.351
##  FINISHED: Min vol ( 0.5 ) reached
```

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

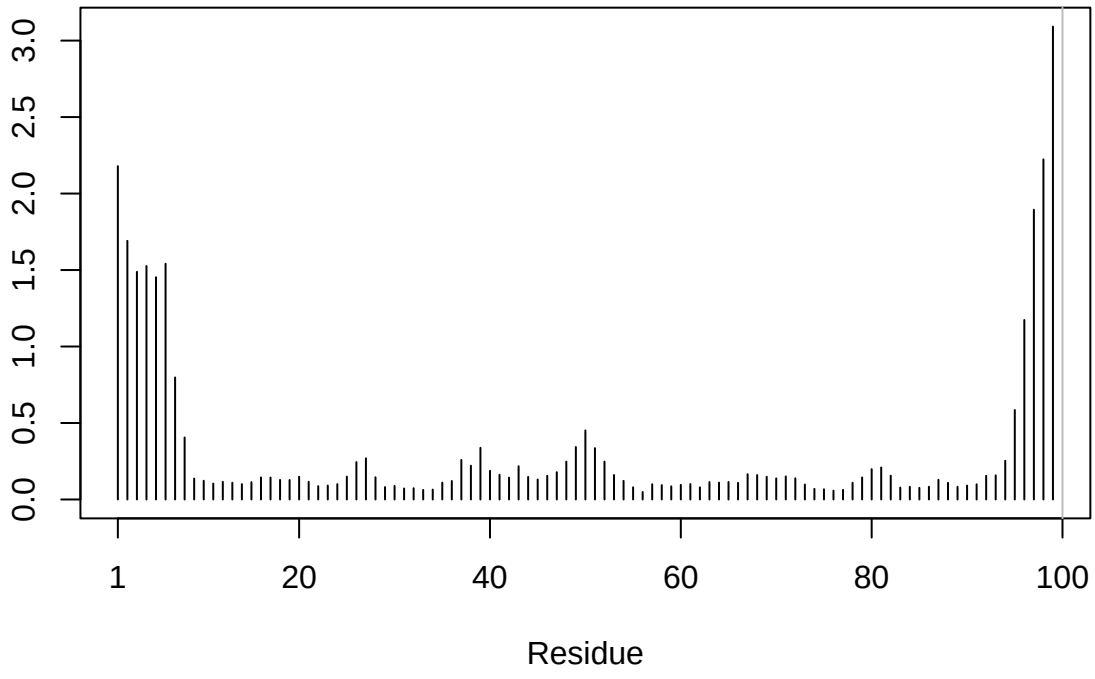
```
## # 91 positions (cumulative volume <= 0.5 Angstrom^3)
##   start end length
## 1     3   3      1
## 2     7  96     90
```

```
xyz <-pdbfit(pdbs, 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 4.5.2
```

```
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"
```

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

```
## [1] 78.12 86.38 87.69 90.12 90.50 93.19
```

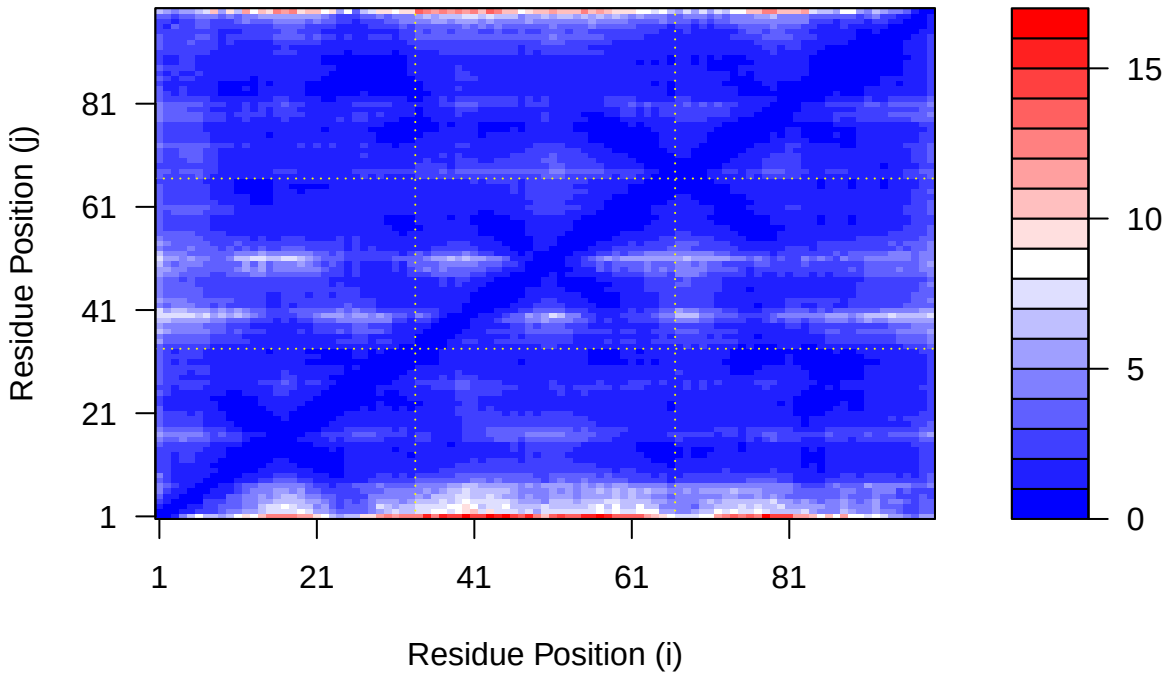
```
pae1$max_pae
```

```
## [1] 17
```

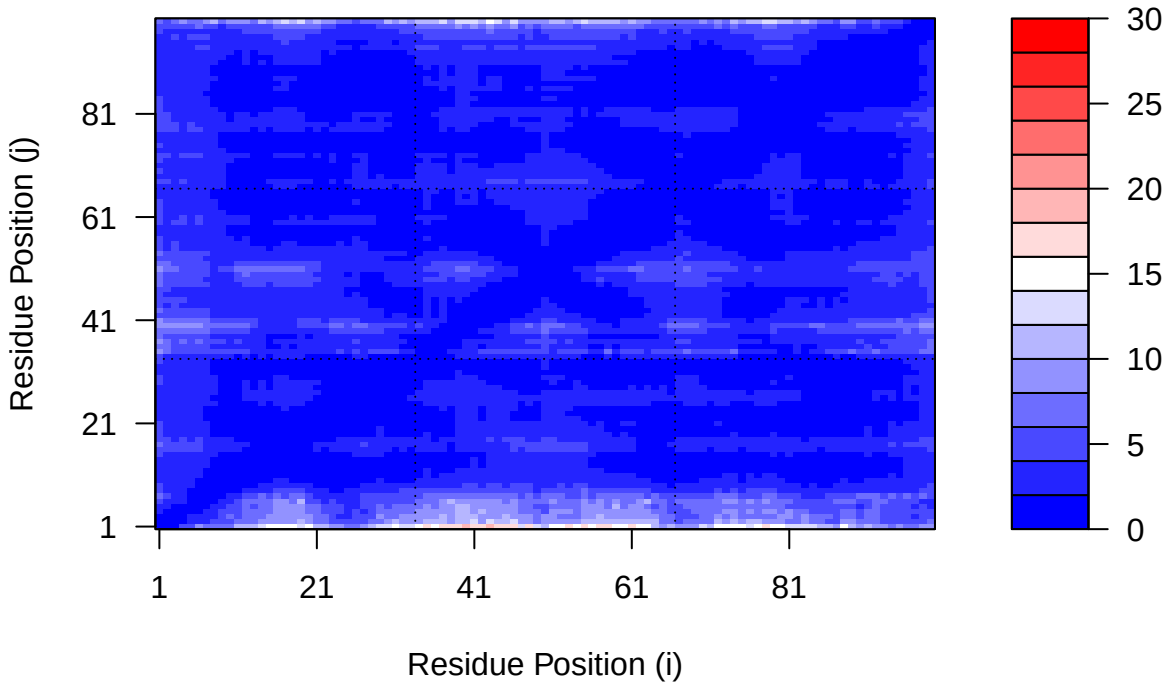
```
pae5$max_pae
```

```
## [1] 19.15625
```

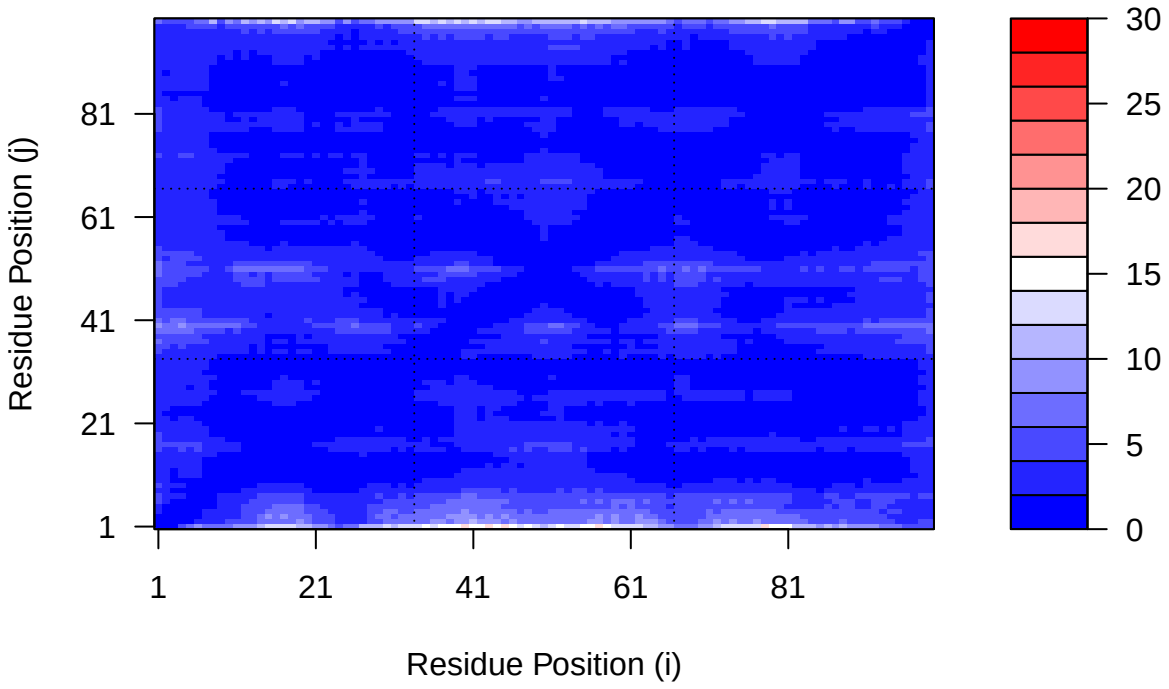
```
plot.dmat(pae1$pae, xlab="Residue Position (i)", ylab="Residue Position (j)")
```




```
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))
```




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

```
## [1] "./Pep_94b5b/Pep_94b5b.a3m"
```

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

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

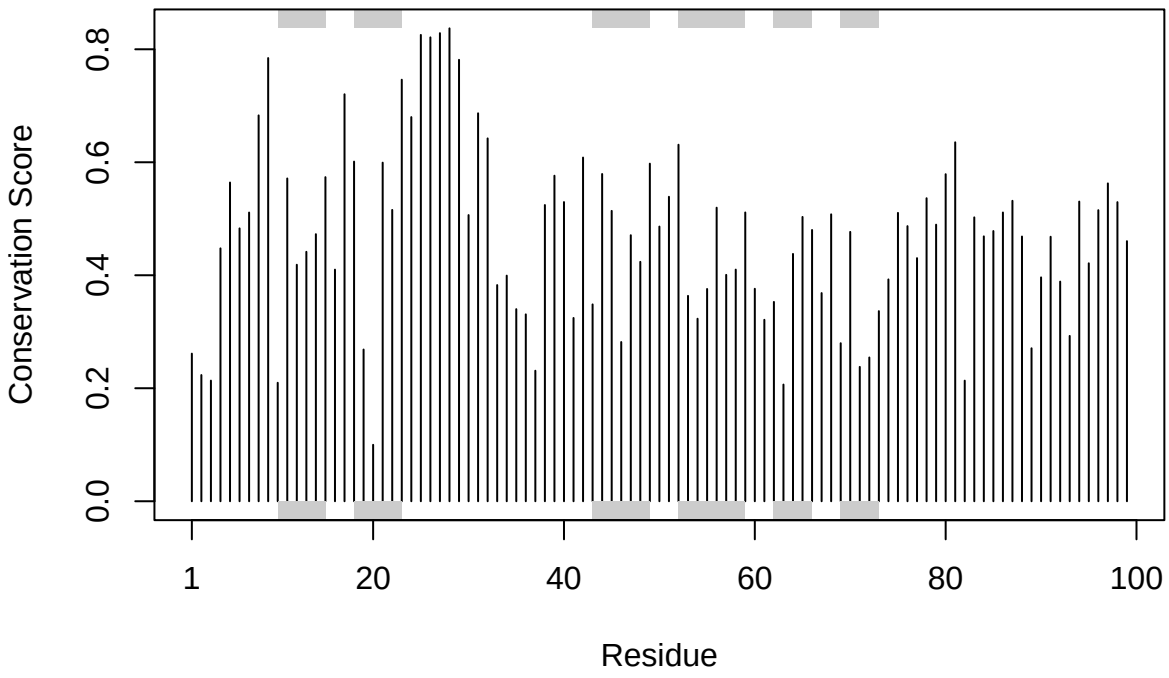
Q. How many sequences are in this alignment?

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

```
## [1] 5397  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("./Pep_94b5b/Pep_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_4_seed_000.pdb")
print(m1.pdb)
```

```
##
##  Call:  read.pdb(file = "./Pep_94b5b/Pep_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_4_seed_000.pdb
##
##    Total Models#: 1
##      Total Atoms#: 757,  XYZs#: 2271  Chains#: 1  (values: A)
##
##      Protein Atoms#: 757  (residues/Calpha atoms#: 99)
##      Nucleic acid Atoms#: 0  (residues/phosphate atoms#: 0)
##
```

```
##      Non-protein/nucleic Atoms#: 0  (residues: 0)
##      Non-protein/nucleic resid values: [ none ]
##
##    Protein sequence:
##       PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYD
##       QILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
##
## + attr: atom, xyz, calpha, call
```

```
m1.pdb$atom$o <- sim[m1.pdb$atom$resno]
write.pdb(m1.pdb, file="m1_conserv.pdb")
```