class11

Muhammad Tariq

2025-05-06

Install packages that were installed in Class10

```
writeLines('PATH="${RTOOLS40_HOME}\\usr\\bin;${PATH}"', con = "~/.Renviron")

pkgbuild::check_build_tools(debug = TRUE)
```

```
## Trying to compile a simple C file
```

```
## Running "C:/PROGRA~1/R/R-4.4.3/bin/x64/Rcmd.exe" SHLIB foo.c
## using C compiler: 'gcc.exe (GCC) 13.3.0'
## gcc  -I"C:/PROGRA~1/R/R-4.4.3/include" -DNDEBUG     -I"C:/rtools44/x86_64-w64-mingw32.static.posix/include"     -O2 -Wall  -mfpmath=sse -msse2 -mstackrealign  -c foo.c -o foo.o
## gcc -shared -s -static-libgcc -o foo.dll tmp.def foo.o -LC:/rtools44/x86_64-w64-mingw32.static.posix/lib/x64 -LC:/rtools44/x86_64-w64-mingw32.static.posix/lib -LC:/PROGRA~1/R/R-4.4.3/bin/x64 -lR
```

```
##
```

```
## Your system is ready to build packages!
```

```
library("tinytex")
library("bio3d")
library("NGLVieweR")
library("pak")

library("BiocManager")
```

```
## Bioconductor version '3.20' is out-of-date; the current release version '3.21'
##   is available with R version '4.5'; see https://bioconductor.org/install
```

```
library("msa")
```

```
## Loading required package: Biostrings
```

```
## Loading required package: BiocGenerics
```

```
##
## Attaching package: 'BiocGenerics'
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
##
##     anyDuplicated, aperm, append, as.data.frame, basename, cbind,
##     colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##     get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
##     match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##     Position, rank, rbind, Reduce, rownames, sapply, saveRDS, setdiff,
##     table, tapply, union, unique, unsplit, which.max, which.min
```

```
## Loading required package: S4Vectors
```

```
## Loading required package: stats4
```

```
##
## Attaching package: 'S4Vectors'
```

```
## The following object is masked from 'package:utils':
##
##     findMatches
```

```
## The following objects are masked from 'package:base':
##
##     expand.grid, I, unname
```

```
## Loading required package: IRanges
```

```
##
## Attaching package: 'IRanges'
```

```
## The following object is masked from 'package:bio3d':
##
##     trim
```

```
## The following object is masked from 'package:grDevices':
##
##     windows
```

```
## Loading required package: XVector
## Loading required package: GenomeInfoDb
```

```
##
## Attaching package: 'Biostrings'
```

```
## The following object is masked from 'package:bio3d':
##
##     mask
```

```
## The following object is masked from 'package:base':
##
##     strsplit
```

```
##
## Attaching package: 'msa'
```

```
## The following object is masked from 'package:BiocManager':
##
##     version
```

```
library("pkgbuild")
```

```
Sys.which("make")
```

```
##                                make
## "C:\\rtools44\\usr\\bin\\make.exe"
```

Load up the pakages we will need for analysis of protein structure sets.

```
library(bio3d)
```

We will analyze the ADK starting with a singe ADK code:“1ake_A”

```
id<- "1ake_A"
aa<- get.seq(id)
```

```
## Warning in get.seq(id): Removing existing file: seqs.fasta
```

```
## Fetching... Please wait. Done.
```

```
aa
```

```
##              1        .         .         .         .         .         60
## pdb|1AKE|A   MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
##              1        .         .         .         .         .         60
##
##             61        .         .         .         .         .         120
## pdb|1AKE|A   DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRI
##             61        .         .         .         .         .         120
##
##            121        .         .         .         .         .         180
## pdb|1AKE|A   VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
##            121        .         .         .         .         .         180
##
##            181        .         .         .   214
## pdb|1AKE|A   YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG
##            181        .         .         .   214
##
## Call:
##   read.fasta(file = outfile)
##
## Class:
##   fasta
##
## Alignment dimensions:
##   1 sequence rows; 214 position columns (214 non-gap, 0 gap)
##
## + attr: id, ali, call
```

Now we can search the PDB database to find all related enteries.

```
blast <- blast.pdb(aa)
```

```
##  Searching ... please wait (updates every 5 seconds) RID = 3JC9SEBV013
##  .................................
##  Reporting 91 hits
```

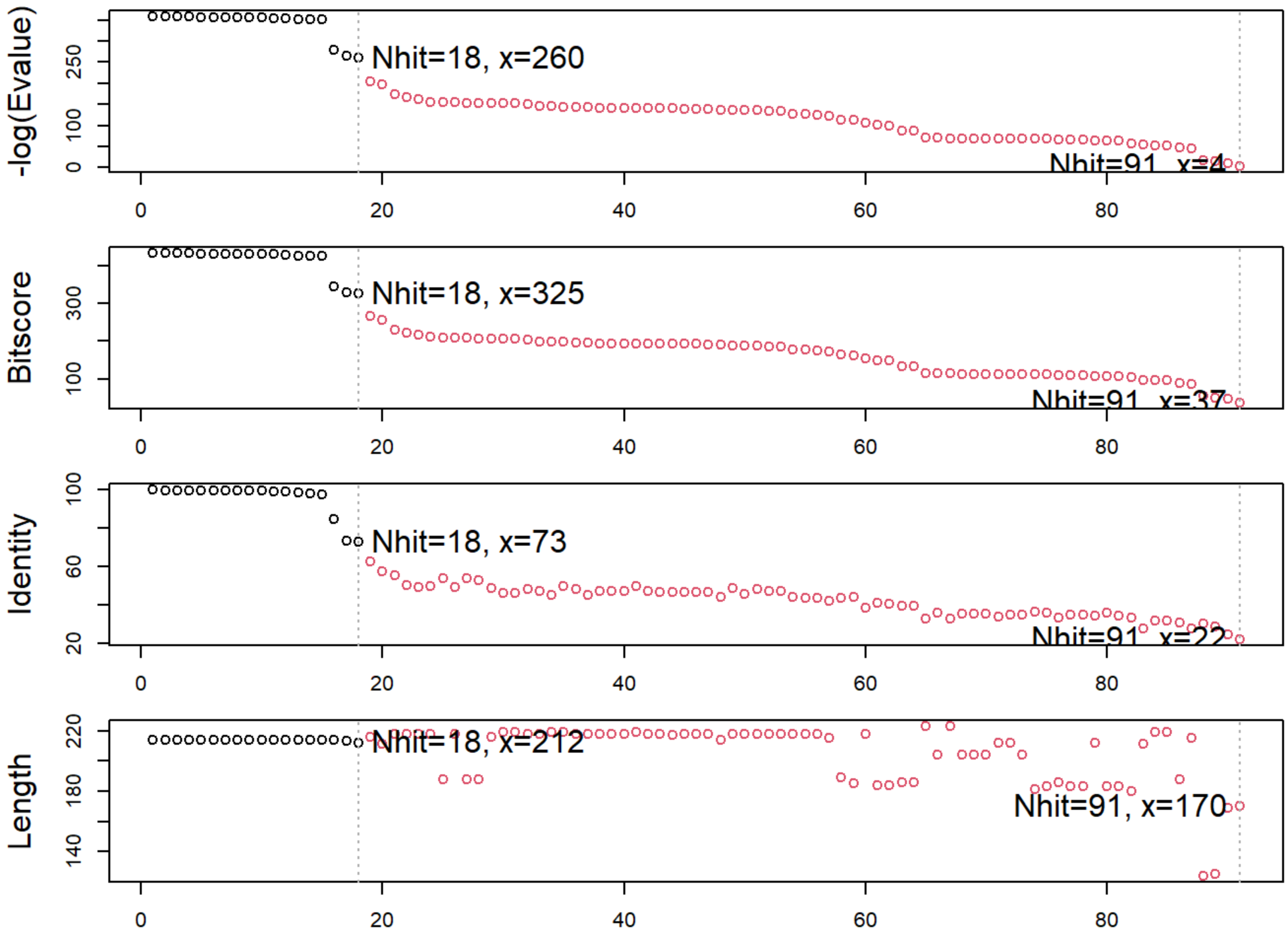
```
attributes(blast)
```

```
## $names
## [1] "hit.tbl" "raw"     "url"
##
## $class
## [1] "blast"
```

Make a little summary figure of these results

```
hits <- plot(blast)
```

```
##   * Possible cutoff values:    260 3
##             Yielding Nhits:    18 91
##
##   * Chosen cutoff value of:    260
##             Yielding Nhits:    18
```



Our “top hits” i.e. the most similar enteries in the database are:

```
hits$pdb.id
```

```
##  [1] "1AKE_A" "8BQF_A" "4X8M_A" "6S36_A" "8Q2B_A" "8RJ9_A" "6RZE_A" "4X8H_A"
##  [9] "3HPR_A" "1E4V_A" "5EJE_A" "1E4Y_A" "3X2S_A" "6HAP_A" "6HAM_A" "8PVW_A"
## [17] "4K46_A" "4NP6_A"
```

```
# Download releated PDB files
files <- get.pdb(hits$pdb.id, path="pdbs", split=TRUE, gzip=TRUE)
```

```
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/1AKE.pdb exists. Skipping download
```

```
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/8BQF.pdb exists. Skipping download
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/4X8M.pdb exists. Skipping download
```

```
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/6S36.pdb exists. Skipping download
```

```
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/8Q2B.pdb exists. Skipping download
```

```
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/8RJ9.pdb exists. Skipping download
```

```
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/6RZE.pdb exists. Skipping download
```

```
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/4X8H.pdb exists. Skipping download
```

```
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/3HPR.pdb exists. Skipping download
```

```
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/1E4V.pdb exists. Skipping download
```

```
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/5EJE.pdb exists. Skipping download
```

```
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/1E4Y.pdb exists. Skipping download
```

```
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/3X2S.pdb exists. Skipping download
```

```
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/6HAP.pdb exists. Skipping download
```

```
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/6HAM.pdb exists. Skipping download
```

```
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/8PVW.pdb exists. Skipping download
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/4K46.pdb exists. Skipping download
```

```
## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
## pdbs/4NP6.pdb exists. Skipping download
```

```
##   |                                                                               |
|   0%  |                                                                               |====
|   6%  |                                                                               |========
|  11%  |                                                                               |========
====                                                          |  17%  |
|================                                                      |  22%  |
|===================                                                   |  28%  |
|=======================                                               |  33%  |
|===========================                                           |  39%  |
|===============================                                       |  44%  |
|===================================                                   |  50%  |
|=======================================                               |  56%  |
|===========================================                           |  61%  |
|===============================================                       |  67%  |
|===================================================                   |  72%  |
|======================================================                |  78%  |
|==========================================================            |  83%  |
|==============================================================        |  89%  |
|==================================================================    |  94%  |
|======================================================================| 100%
```

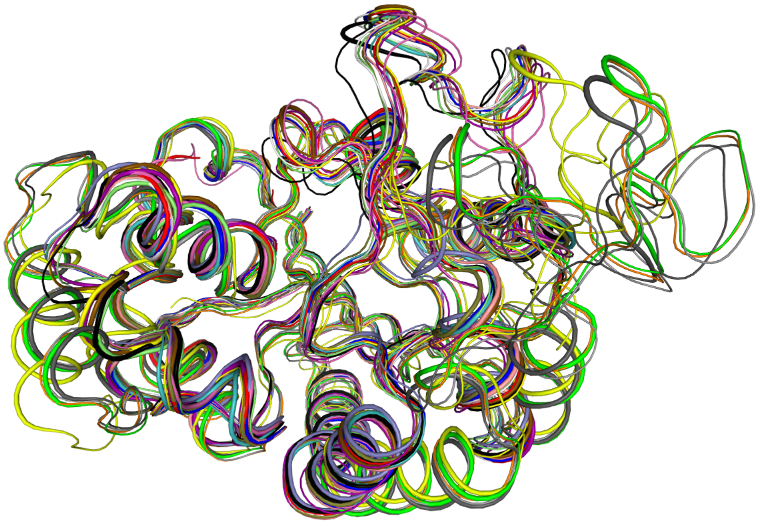
Align and supperpose these structures

```
# Align releated PDBs
pdbs <- pdbaln(files, fit = TRUE, exefile="msa")
```

```
## Reading PDB files:
## pdbs/split_chain/1AKE_A.pdb
## pdbs/split_chain/8BQF_A.pdb
## pdbs/split_chain/4X8M_A.pdb
## pdbs/split_chain/6S36_A.pdb
## pdbs/split_chain/8Q2B_A.pdb
## pdbs/split_chain/8RJ9_A.pdb
## pdbs/split_chain/6RZE_A.pdb
## pdbs/split_chain/4X8H_A.pdb
## pdbs/split_chain/3HPR_A.pdb
## pdbs/split_chain/1E4V_A.pdb
## pdbs/split_chain/5EJE_A.pdb
## pdbs/split_chain/1E4Y_A.pdb
## pdbs/split_chain/3X2S_A.pdb
## pdbs/split_chain/6HAP_A.pdb
## pdbs/split_chain/6HAM_A.pdb
## pdbs/split_chain/8PVW_A.pdb
## pdbs/split_chain/4K46_A.pdb
## pdbs/split_chain/4NP6_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## .   PDB has ALT records, taking A only, rm.alt=TRUE
## ..   PDB has ALT records, taking A only, rm.alt=TRUE
## .   PDB has ALT records, taking A only, rm.alt=TRUE
## .   PDB has ALT records, taking A only, rm.alt=TRUE
## .   PDB has ALT records, taking A only, rm.alt=TRUE
## ..   PDB has ALT records, taking A only, rm.alt=TRUE
## ..   PDB has ALT records, taking A only, rm.alt=TRUE
## ....   PDB has ALT records, taking A only, rm.alt=TRUE
## .   PDB has ALT records, taking A only, rm.alt=TRUE
## .   PDB has ALT records, taking A only, rm.alt=TRUE
## ..
##
## Extracting sequences
##
## pdb/seq: 1   name: pdbs/split_chain/1AKE_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 2   name: pdbs/split_chain/8BQF_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 3   name: pdbs/split_chain/4X8M_A.pdb
## pdb/seq: 4   name: pdbs/split_chain/6S36_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 5   name: pdbs/split_chain/8Q2B_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 6   name: pdbs/split_chain/8RJ9_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 7   name: pdbs/split_chain/6RZE_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 8   name: pdbs/split_chain/4X8H_A.pdb
## pdb/seq: 9   name: pdbs/split_chain/3HPR_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 10   name: pdbs/split_chain/1E4V_A.pdb
## pdb/seq: 11   name: pdbs/split_chain/5EJE_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 12   name: pdbs/split_chain/1E4Y_A.pdb
## pdb/seq: 13   name: pdbs/split_chain/3X2S_A.pdb
## pdb/seq: 14   name: pdbs/split_chain/6HAP_A.pdb
## pdb/seq: 15   name: pdbs/split_chain/6HAM_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 16   name: pdbs/split_chain/8PVW_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 17   name: pdbs/split_chain/4K46_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 18   name: pdbs/split_chain/4NP6_A.pdb
```

Side-note:

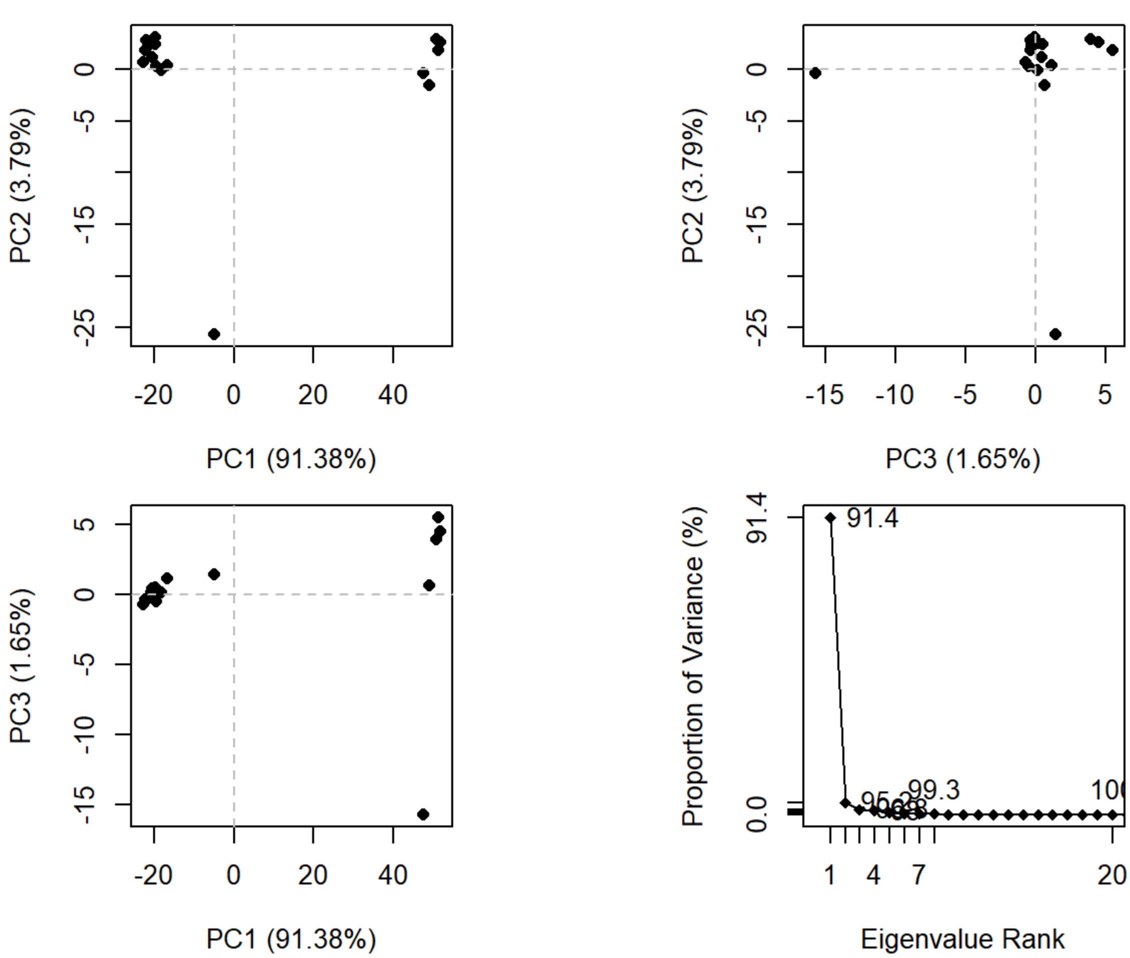
```
library(bio3dview)

view.pdbs(pdbs)
```



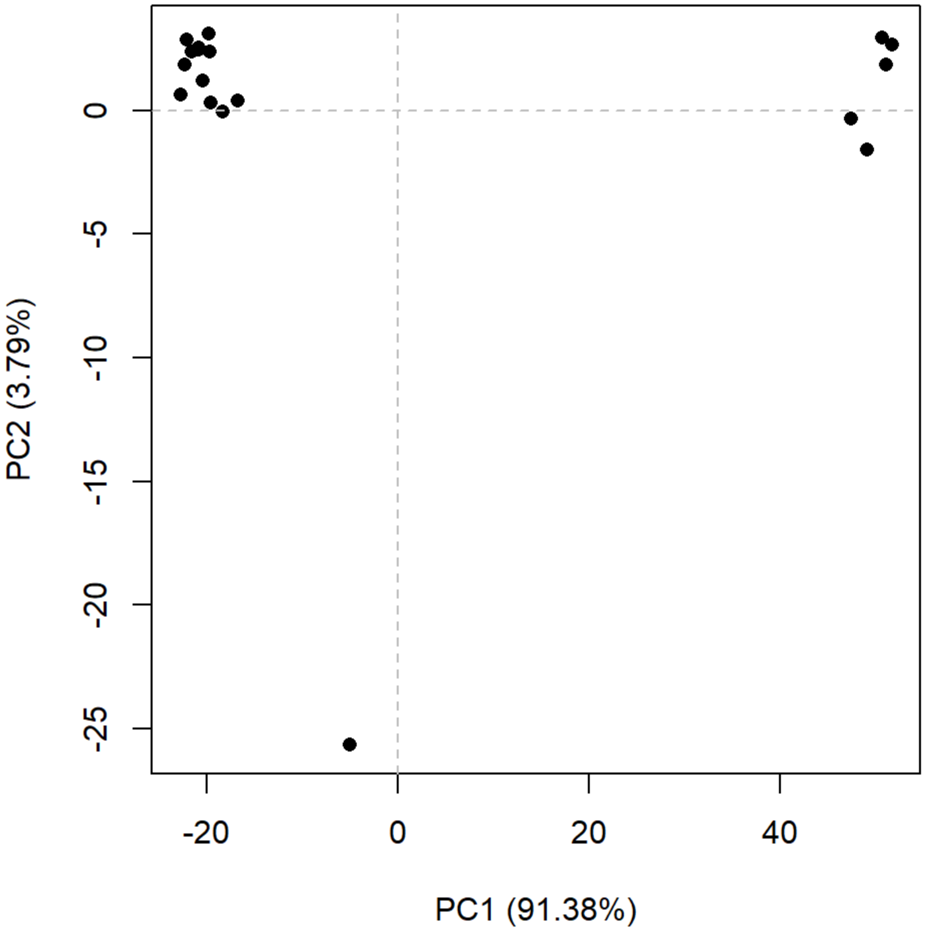
This is better but still difficult to see what is similar and diffeent in all these structures or indeed learn much about how this family works.

Lets try PCA:

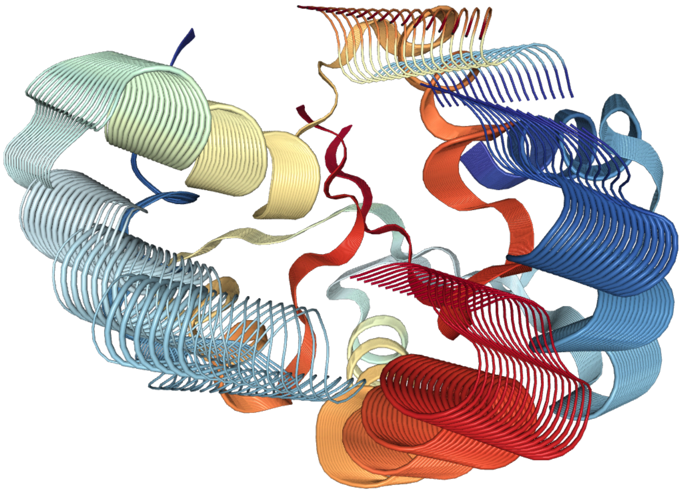
```
pc <- pca(pdbs)
plot(pc)
```



```
plot(pc, pc.axes=1:2)
```



```
view.pca(pc)
```



write a PDB “trajectory” for mol-star

```
mktrj(pc, file="pca_results.pdb")
```