Class 09: Structural Bioinformatics 1

Lawrence Adhinatha

2023-02-07

What is in the PDB anyway?

Tha main database of biomolecular structures is called the PDB and is available at www.rcsb.org.

Let's begin by seeing what is in this database:

```
stats <- read.csv("Data Export Summary.csv", row.names=1)
stats_df <- as.data.frame(stats)
stats_df
```

```
##                            X.ray    EM    NMR Multiple.methods Neutron Other
## Protein (only)           152,809 9,421 12,117               191      72    32
## Protein/Oligosaccharide    9,008 1,654     32                 7       1     0
## Protein/NA                 8,061 2,944    281                 6       0     0
## Nucleic acid (only)        2,602    77  1,433                12       2     1
## Other                        163     9     31                 0       0     0
## Oligosaccharide (only)        11     0      6                 1       0     4
##                            Total
## Protein (only)           174,642
## Protein/Oligosaccharide   10,702
## Protein/NA                11,292
## Nucleic acid (only)        4,127
## Other                        203
## Oligosaccharide (only)        22
```

Q1: What percentage of structures in the PDB are solved by X-Ray and Electron Microscopy.

```
# delete commas to make values numeric
n.xray <- sum( as.numeric(gsub(",", "", stats_df[,1])))
n.em <- sum( as.numeric(gsub(",", "", stats_df[,2])))
n.total <- sum( as.numeric(gsub(",", "", stats_df[,7])))

# calculate percent to 2 sigfigs
p.stats <- function(num, tot) {
  percent <- round(((num/tot)*100), 2)
  percent
}

p.stats(n.xray, n.total)
```

```
## [1] 85.9
```

```
p.stats(n.em, n.total)
```

```
## [1] 7.02
```

There are `p.stats(n.xray, n.total%` xray structures and `p.stats(n.em, n.total)%` EM structures in the current PDB database.

> Q2: What proportion of structures in the PDB are protein?

```
n.prot <- sum(as.numeric(gsub(",", "", stats_df[1:3,7])))
p.prot <- p.stats(n.prot, n.total)
p.prot
```

```
## [1] 97.83
```

```
ind.prot <- (as.numeric(gsub(",", "", stats_df[1:3,7])))
ind.prot/n.total
```

```
## [1] 0.86891755 0.05324696 0.05618246
```

`ind.prot/n.total` of the structures in the PDB are protein.

> Q3: Type HIV in the PDB website search box on the home page and determine how many HIV-1 protease structures are in the current PDB?

It is not straightforward to find all HIV-1 protease structures using plain text searching in the database.

Visualizing the HIV-1 protease structure

A wee pic of HIV-1 Protease from Molstar.

> Q4: Water molecules normally have 3 atoms. Why do we see just one atom per water molecule in this structure?

The resolution of the structure file is 2 Angstroms. A Hydrogen atom is about 1 Angstrom in size, requiring a resolution smaller than that to view it. The atom shown is the Oxygen in the water molecule.

> Q5: There is a critical "conserved" water molecule in the binding site. Can you identify this water molecule? What residue number does this water molecule have

HOH 308

> Q6: Generate and save a figure clearly showing the two distinct chains of HIV-protease along with the ligand. You might also consider showing the catalytic residues ASP 25 in each chain and the critical water (we recommend "Ball & Stick" for these side-chains). Add this figure to your Quarto document.

Introduction to Bio3D in R

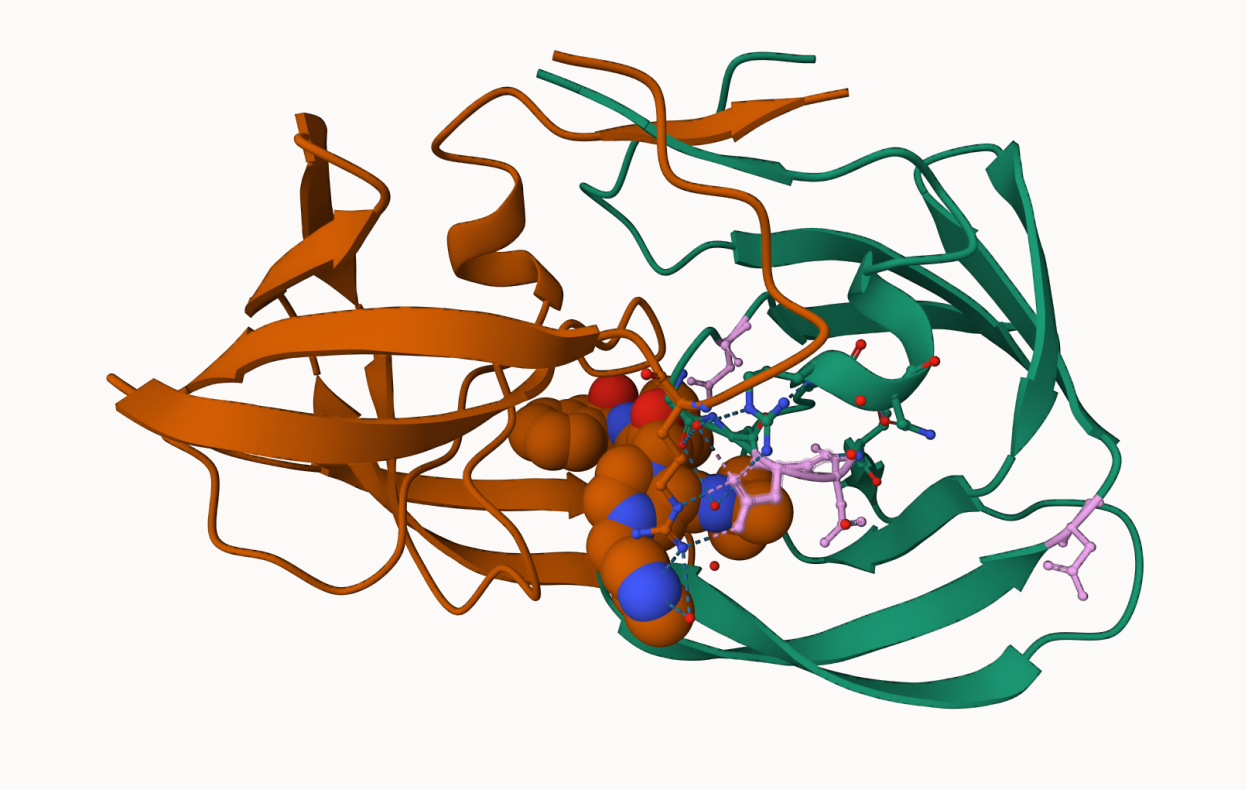


Figure 1: An image obtained from Molstar sufficiently worthy of display for this assignment.

```
# Use Bio3D
library(bio3d)
pdb <- read.pdb("1hsg")
```

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

```
pdb
```

```
##
##  Call:  read.pdb(file = "1hsg")
##
##    Total Models#: 1
##      Total Atoms#: 1686,  XYZs#: 5058  Chains#: 2  (values: A B)
##
##      Protein Atoms#: 1514  (residues/Calpha atoms#: 198)
##      Nucleic acid Atoms#: 0  (residues/phosphate atoms#: 0)
##
##      Non-protein/nucleic Atoms#: 172  (residues: 128)
##      Non-protein/nucleic resid values: [ HOH (127), MK1 (1) ]
##
##    Protein sequence:
##       PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYD
##       QILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPQITLWQRPLVTIKIGGQLKE
##       ALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTP
```

```
##       VNIIGRNLLTQIGCTLNF
##
## + attr: atom, xyz, seqres, helix, sheet,
##         calpha, remark, call
```

```
head(pdb$atom)
```

```
##   type eleno elety  alt resid chain resno insert      x      y     z o     b
## 1 ATOM     1     N <NA>   PRO     A     1   <NA> 29.361 39.686 5.862 1 38.10
## 2 ATOM     2    CA <NA>   PRO     A     1   <NA> 30.307 38.663 5.319 1 40.62
## 3 ATOM     3     C <NA>   PRO     A     1   <NA> 29.760 38.071 4.022 1 42.64
## 4 ATOM     4     O <NA>   PRO     A     1   <NA> 28.600 38.302 3.676 1 43.40
## 5 ATOM     5    CB <NA>   PRO     A     1   <NA> 30.508 37.541 6.342 1 37.87
## 6 ATOM     6    CG <NA>   PRO     A     1   <NA> 29.296 37.591 7.162 1 38.40
##   segid elesy charge
## 1  <NA>     N   <NA>
## 2  <NA>     C   <NA>
## 3  <NA>     C   <NA>
## 4  <NA>     O   <NA>
## 5  <NA>     C   <NA>
## 6  <NA>     C   <NA>
```

What is the first residue 3 letter code?

```
pdb$atom$resid[1]
```

```
## [1] "PRO"
```

```
aa321(pdb$atom$resid[1])
```

```
## [1] "P"
```

Q7: How many amino acid residues are there in this pdb object?

198 aa residues

Q8: Name one of the two non-protein residues?

HOH

Q9: How many protein chains are in this structure?

2

```
attributes(pdb)
```

```
## $names
## [1] "atom"   "xyz"    "seqres" "helix"  "sheet"  "calpha" "remark" "call"
##
## $class
## [1] "pdb" "sse"
```

Predicting functional motions of a single structure

Let's read a new PDB structure of Adenylate Kinase (PDB code: 6s36) and perform Normal mode analysis.

```
adk <- read.pdb("6s36")
```

```
##   Note: Accessing on-line PDB file
##    PDB has ALT records, taking A only, rm.alt=TRUE
```

```
adk
```

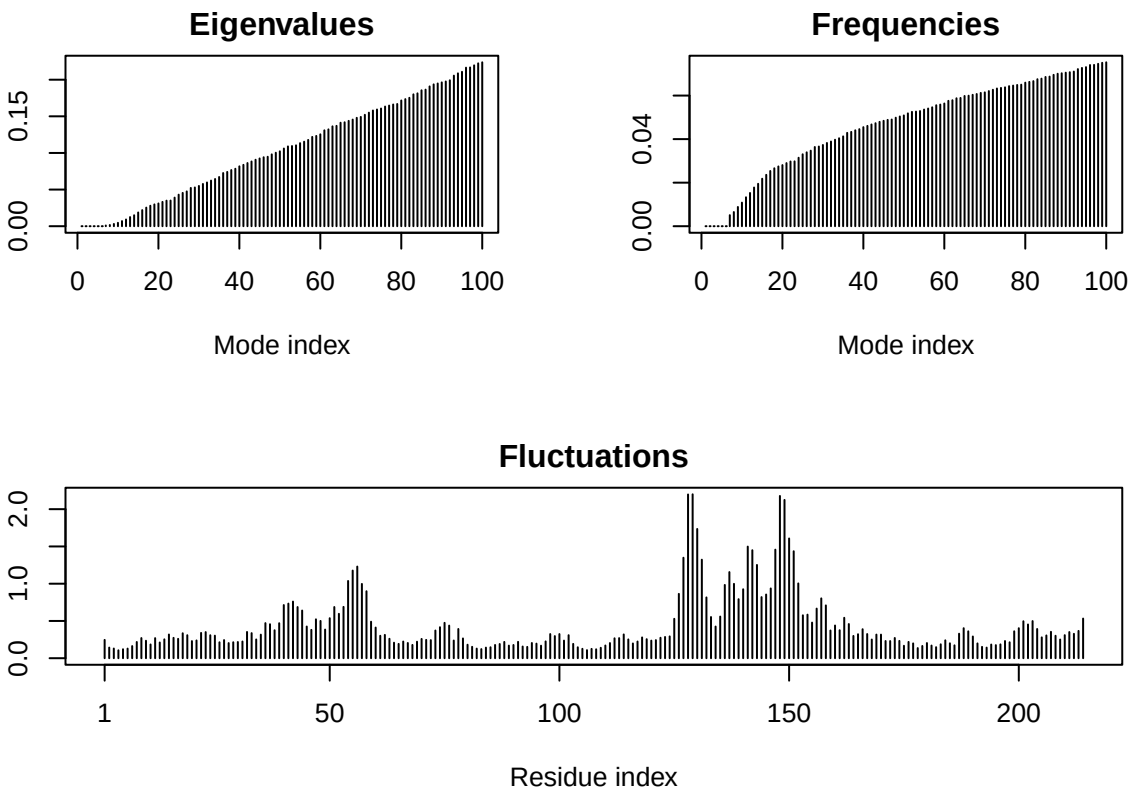
```
##
##  Call:  read.pdb(file = "6s36")
##
##    Total Models#: 1
##      Total Atoms#: 1898,  XYZs#: 5694  Chains#: 1  (values: A)
##
##      Protein Atoms#: 1654  (residues/Calpha atoms#: 214)
##      Nucleic acid Atoms#: 0  (residues/phosphate atoms#: 0)
##
##      Non-protein/nucleic Atoms#: 244  (residues: 244)
##      Non-protein/nucleic resid values: [ CL (3), HOH (238), MG (2), NA (1) ]
##
##    Protein sequence:
##       MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
##       DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDKI
##       VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
##       YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG
##
## + attr: atom, xyz, seqres, helix, sheet,
##         calpha, remark, call
```

Normal mode analysis (NMA) is a structural bioinformatics method to predict protein flexibility and potential functional motions (a.k.a. conformational changes).

```
m <- nma(adk)
```

```
##  Building Hessian...     Done in 0.022 seconds.
##  Diagonalizing Hessian...    Done in 0.449 seconds.
```

```
plot(m)
```




```
mktrj(m, file="adk_m7.pdb")
```

Comparative structure analysis of Adenylate Kinase

We will now be conducting a Principal Component Analysis on the complete collection of Adenylate kinase structures in the PDB.

We begin with getting a singl eprotein sequence for a protein family of interest.

Q10. Which of the packages above is found only on BioConductor and not CRAN?

msa package.

Q11. Which of the above packages is not found on BioConductor or CRAN?:

bio3d view

Q12. True or False? Functions from the devtools package can be used to install packages from GitHub and BitBucket?

True

```
library(bio3d)

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

```
## Warning in get.seq("1ake_A"): 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
```

Q13. How many amino acids are in this sequence, i.e. how long is this sequence?

214 amino acids long

Now we can use this sequence as a query to BLAST search the PDB to find similar sequences and structures.

```
# Blast or hmmer search
# b <- blast.pdb(aa)
```

Rather than re-running the BLAST search every time this document loads (and save these results for this point in time), we can save and load the blast results.

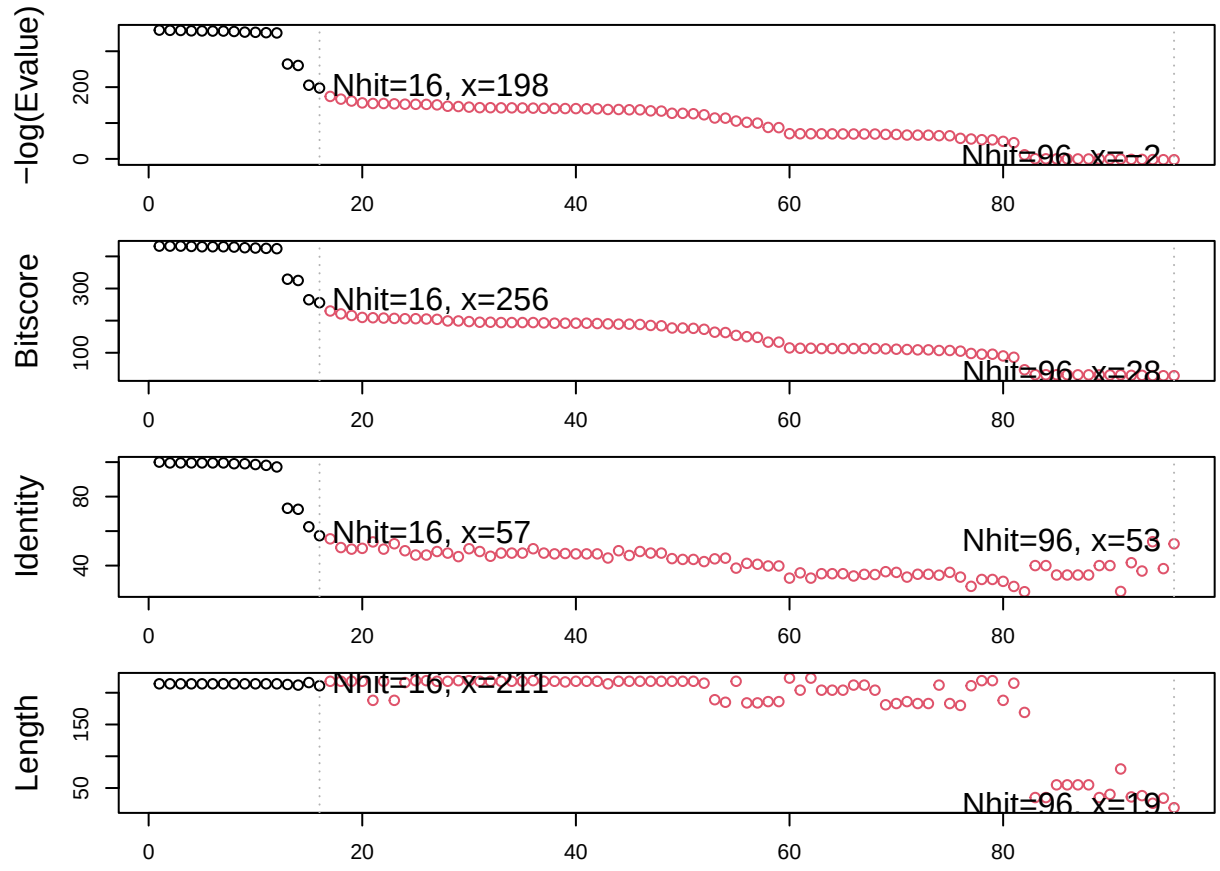
```
# saveRDS(b, file="blast_results.RDS")
```

```
b <- readRDS("blast_results.RDS")
```

A summary plot of our BLAST results

```
hits <- plot(b)
```

```
##   * Possible cutoff values:    197 -3
##              Yielding Nhits:    16 96
##
##   * Chosen cutoff value of:    197
##              Yielding Nhits:    16
```




```
hits
```

```
## $hits
##    pdb.id   acc      group
## 1  "1AKE_A" "1AKE_A" "1"
## 2  "4X8M_A" "4X8M_A" "1"
## 3  "6S36_A" "6S36_A" "1"
## 4  "6RZE_A" "6RZE_A" "1"
## 5  "4X8H_A" "4X8H_A" "1"
## 6  "3HPR_A" "3HPR_A" "1"
## 7  "1E4V_A" "1E4V_A" "1"
## 8  "5EJE_A" "5EJE_A" "1"
```

```
## 9  "1E4Y_A" "1E4Y_A" "1"
## 10 "3X2S_A" "3X2S_A" "1"
## 11 "6HAP_A" "6HAP_A" "1"
## 12 "6HAM_A" "6HAM_A" "1"
## 13 "4K46_A" "4K46_A" "1"
## 14 "4NP6_A" "4NP6_A" "1"
## 15 "3GMT_A" "3GMT_A" "1"
## 16 "4PZL_A" "4PZL_A" "1"
##
## $pdb.id
##  [1] "1AKE_A" "4X8M_A" "6S36_A" "6RZE_A" "4X8H_A" "3HPR_A" "1E4V_A" "5EJE_A"
##  [9] "1E4Y_A" "3X2S_A" "6HAP_A" "6HAM_A" "4K46_A" "4NP6_A" "3GMT_A" "4PZL_A"
##
## $acc
##  [1] "1AKE_A" "4X8M_A" "6S36_A" "6RZE_A" "4X8H_A" "3HPR_A" "1E4V_A" "5EJE_A"
##  [9] "1E4Y_A" "3X2S_A" "6HAP_A" "6HAM_A" "4K46_A" "4NP6_A" "3GMT_A" "4PZL_A"
##
## $inds
##  [1]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [13]  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##
## attr(,"class")
## [1] "blast"
```

```
hits$pdb.id
```

```
##  [1] "1AKE_A" "4X8M_A" "6S36_A" "6RZE_A" "4X8H_A" "3HPR_A" "1E4V_A" "5EJE_A"
##  [9] "1E4Y_A" "3X2S_A" "6HAP_A" "6HAM_A" "4K46_A" "4NP6_A" "3GMT_A" "4PZL_A"
```

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Align and superpose structures!

We will now use the `pdbaln()` function to optionally fit (i.e. superpose) the identified PDB structures.

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

```
## Reading PDB files:
## pdbs/1AKE.pdb
## pdbs/4X8M.pdb
## pdbs/6S36.pdb
## pdbs/6RZE.pdb
## pdbs/4X8H.pdb
## pdbs/3HPR.pdb
## pdbs/1E4V.pdb
## pdbs/5EJE.pdb
## pdbs/1E4Y.pdb
```

```
## pdbs/3X2S.pdb
## pdbs/6HAP.pdb
## pdbs/6HAM.pdb
## pdbs/4K46.pdb
## pdbs/4NP6.pdb
## pdbs/3GMT.pdb
## pdbs/4PZL.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
## .
##
## Extracting sequences
##
## pdb/seq: 1   name: pdbs/1AKE.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 2   name: pdbs/4X8M.pdb
## pdb/seq: 3   name: pdbs/6S36.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 4   name: pdbs/6RZE.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 5   name: pdbs/4X8H.pdb
## pdb/seq: 6   name: pdbs/3HPR.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 7   name: pdbs/1E4V.pdb
## pdb/seq: 8   name: pdbs/5EJE.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 9   name: pdbs/1E4Y.pdb
## pdb/seq: 10   name: pdbs/3X2S.pdb
## pdb/seq: 11   name: pdbs/6HAP.pdb
## pdb/seq: 12   name: pdbs/6HAM.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 13   name: pdbs/4K46.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 14   name: pdbs/4NP6.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 15   name: pdbs/3GMT.pdb
## pdb/seq: 16   name: pdbs/4PZL.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
```

```
pdbs
```

```
##                1        .        .        .        .        .         60
## pdbs/1AKE.pdb   ------------------------------------------------------------
## pdbs/4X8M.pdb   ------------------------------------------------------------
## pdbs/6S36.pdb   ------------------------------------------------------------
## pdbs/6RZE.pdb   ------------------------------------------------------------
## pdbs/4X8H.pdb   ------------------------------------------------------------
```

```
## pdbs/3HPR.pdb   ------------------------------------------------------------
## pdbs/1E4V.pdb   ------------------------------------------------------------
## pdbs/5EJE.pdb   ------------------------------------------------------------
## pdbs/1E4Y.pdb   ------------------------------------------------------------
## pdbs/3X2S.pdb   ------------------------------------------------------------
## pdbs/6HAP.pdb   ------------------------------------------------------------
## pdbs/6HAM.pdb   ------------------------------------------------------------
## pdbs/4K46.pdb   ------------------------------------------------------------
## pdbs/4NP6.pdb   --------NAMRIILLGAPGAGKGTQAQFIMEKFGIPQISTGDMLRAAIKAGTELGKQAK
## pdbs/3GMT.pdb   ------------------------------------------------------------
## pdbs/4PZL.pdb   TENLYFQSNAMRIILLGAPGAGKGTQAKIIEQKYNIAHISTGDMIRETIKSGSALGQELK
##
##                  1        .         .         .         .         .         60
##
##                 61        .         .         .         .         .         120
## pdbs/1AKE.pdb   ------------------------------------------------------------
## pdbs/4X8M.pdb   ------------------------------------------------------------
## pdbs/6S36.pdb   ------------------------------------------------------------
## pdbs/6RZE.pdb   ------------------------------------------------------------
## pdbs/4X8H.pdb   ------------------------------------------------------------
## pdbs/3HPR.pdb   ------------------------------------------------------------
## pdbs/1E4V.pdb   ------------------------------------------------------------
## pdbs/5EJE.pdb   ------------------------------------------------------------
## pdbs/1E4Y.pdb   ------------------------------------------------------------
## pdbs/3X2S.pdb   ------------------------------------------------------------
## pdbs/6HAP.pdb   ------------------------------------------------------------
## pdbs/6HAM.pdb   ------------------------------------------------------------
## pdbs/4K46.pdb   ------------------------------------------------------------
## pdbs/4NP6.pdb   AVIDAGQLVSDDIILGLIKERIAQADCEKGFLLDGFPRTIPQADGLKEMGINVDYVIEFD
## pdbs/3GMT.pdb   ------------------------------------------------------------
## pdbs/4PZL.pdb   KVLDAGELVSDEFIIKIVKDRISKNDCNNGFLLDGVPRTIPQAQELDKLGVNIDYIVEVD
##
##                 61        .         .         .         .         .         120
##
##                121        .         .         .         .         .         180
## pdbs/1AKE.pdb   ------------------------------------------------------------
## pdbs/4X8M.pdb   ------------------------------------------------------------
## pdbs/6S36.pdb   ------------------------------------------------------------
## pdbs/6RZE.pdb   ------------------------------------------------------------
## pdbs/4X8H.pdb   ------------------------------------------------------------
## pdbs/3HPR.pdb   ------------------------------------------------------------
## pdbs/1E4V.pdb   ------------------------------------------------------------
## pdbs/5EJE.pdb   ------------------------------------------------------------
## pdbs/1E4Y.pdb   ------------------------------------------------------------
## pdbs/3X2S.pdb   ------------------------------------------------------------
## pdbs/6HAP.pdb   ------------------------------------------------------------
## pdbs/6HAM.pdb   ------------------------------------------------------------
## pdbs/4K46.pdb   ------------------------------------------------------------
## pdbs/4NP6.pdb   VADDVIVERMAGRRAHLPSGRTYHVVYNPPKVEGKDDVTGEDLVIREDDKEETVRARLNV
## pdbs/3GMT.pdb   ------------------------------------------------------------
## pdbs/4PZL.pdb   VADNLLIERITGRRIHPASGRTYHTKFNPPKVADKDDVTGEPLITRTDDNEDTVKQRLSV
##
##                121        .         .         .         .         .         180
##
```

```
##               181        .         .         .         .         .         240
## pdbs/1AKE.pdb   ------------------------------------------------------------
## pdbs/4X8M.pdb   ------------------------------------------------------------
## pdbs/6S36.pdb   ------------------------------------------------------------
## pdbs/6RZE.pdb   ------------------------------------------------------------
## pdbs/4X8H.pdb   ------------------------------------------------------------
## pdbs/3HPR.pdb   ------------------------------------------------------------
## pdbs/1E4V.pdb   ------------------------------------------------------------
## pdbs/5EJE.pdb   ------------------------------------------------------------
## pdbs/1E4Y.pdb   ------------------------------------------------------------
## pdbs/3X2S.pdb   ------------------------------------------------------------
## pdbs/6HAP.pdb   ------------------------------------------------------------
## pdbs/6HAM.pdb   ------------------------------------------------------------
## pdbs/4K46.pdb   ------------------------------------------------------------
## pdbs/4NP6.pdb   YHTQTAPLIEYYGK-----------EAAAGKTQYLKFDGTKQVSEVSADIAKALAAMRI
## pdbs/3GMT.pdb   ------------------------------------------------------------
## pdbs/4PZL.pdb   YHAQTAKLIDFYRNFSSTNTKIPKYIKINGDQAVEKVSQDIFDQLNKTENLYFQSNAMRI
##
##               181        .         .         .         .         .         240
##
##               241        .         .         .         .         .         300
## pdbs/1AKE.pdb   ------------------------------------------------------------
## pdbs/4X8M.pdb   ------------------------------------------------------------
## pdbs/6S36.pdb   ------------------------------------------------------------
## pdbs/6RZE.pdb   ------------------------------------------------------------
## pdbs/4X8H.pdb   ------------------------------------------------------------
## pdbs/3HPR.pdb   ------------------------------------------------------------
## pdbs/1E4V.pdb   ------------------------------------------------------------
## pdbs/5EJE.pdb   ------------------------------------------------------------
## pdbs/1E4Y.pdb   ------------------------------------------------------------
## pdbs/3X2S.pdb   ------------------------------------------------------------
## pdbs/6HAP.pdb   ------------------------------------------------------------
## pdbs/6HAM.pdb   ------------------------------------------------------------
## pdbs/4K46.pdb   ------------------------------------------------------------
## pdbs/4NP6.pdb   ILLGAPGAGKGTQAQFIMEKFGIPQISTGDMLRAAIKAGTELGKQAKAVIDAGQLVSDDI
## pdbs/3GMT.pdb   ------------------------------------------------------------
## pdbs/4PZL.pdb   ILLGAPGAGKGTQAKIIEQKYNIAHISTGDMIRETIKSGSALGQELKKVLDAGELVSDEF
##
##               241        .         .         .         .         .         300
##
##               301        .         .         .         .         .         360
## pdbs/1AKE.pdb   ------------------------------------------------------------
## pdbs/4X8M.pdb   ------------------------------------------------------------
## pdbs/6S36.pdb   ------------------------------------------------------------
## pdbs/6RZE.pdb   ------------------------------------------------------------
## pdbs/4X8H.pdb   ------------------------------------------------------------
## pdbs/3HPR.pdb   ------------------------------------------------------------
## pdbs/1E4V.pdb   ------------------------------------------------------------
## pdbs/5EJE.pdb   ------------------------------------------------------------
## pdbs/1E4Y.pdb   ------------------------------------------------------------
## pdbs/3X2S.pdb   ------------------------------------------------------------
## pdbs/6HAP.pdb   ------------------------------------------------------------
## pdbs/6HAM.pdb   ------------------------------------------------------------
## pdbs/4K46.pdb   ------------------------------------------------------------
```

```
## pdbs/4NP6.pdb   ILGLIKERIAQADCEKGFLLDGFPRTIPQADGLKEMGINVDYVIEFDVADDVIVERMAGR
## pdbs/3GMT.pdb   ------------------------------------------------------------
## pdbs/4PZL.pdb   IIKIVKDRISKNDCNNGFLLDGVPRTIPQAQELDKLGVNIDYIVEVDVADNLLIERITGR
##
##               301        .         .         .         .         .         360
##
##               361        .         .         .         .         .         420
## pdbs/1AKE.pdb   ------------------------------------------------------------
## pdbs/4X8M.pdb   ------------------------------------------------------------
## pdbs/6S36.pdb   ------------------------------------------------------------
## pdbs/6RZE.pdb   ------------------------------------------------------------
## pdbs/4X8H.pdb   ------------------------------------------------------------
## pdbs/3HPR.pdb   ------------------------------------------------------------
## pdbs/1E4V.pdb   ------------------------------------------------------------
## pdbs/5EJE.pdb   ------------------------------------------------------------
## pdbs/1E4Y.pdb   ------------------------------------------------------------
## pdbs/3X2S.pdb   ------------------------------------------------------------
## pdbs/6HAP.pdb   ------------------------------------------------------------
## pdbs/6HAM.pdb   ------------------------------------------------------------
## pdbs/4K46.pdb   ------------------------------------------------------------
## pdbs/4NP6.pdb   RAHLPSGRTYHVVYNPPKVEGKDDVTGEDLVIREDDKEETVRARLNVYHTQTAPLIEYYG
## pdbs/3GMT.pdb   ------------------------------------------------------------
## pdbs/4PZL.pdb   RIHPASGRTYHTKFNPPKVADKDDVTGEPLITRTDDNEDTVKQRLSVYHAQTAKLIDFYR
##
##               361        .         .         .         .         .         420
##
##               421        .         .         .         .         .         480
## pdbs/1AKE.pdb   ---------------------------------------MRIILLGAPGAGKGTQAQFIMEK
## pdbs/4X8M.pdb   ---------------------------------------MRIILLGAPGAGKGTQAQFIMEK
## pdbs/6S36.pdb   ---------------------------------------MRIILLGAPGAGKGTQAQFIMEK
## pdbs/6RZE.pdb   ---------------------------------------MRIILLGAPGAGKGTQAQFIMEK
## pdbs/4X8H.pdb   ---------------------------------------MRIILLGAPGAGKGTQAQFIMEK
## pdbs/3HPR.pdb   ---------------------------------------MRIILLGAPGAGKGTQAQFIMEK
## pdbs/1E4V.pdb   ---------------------------------------MRIILLGAPVAGKGTQAQFIMEK
## pdbs/5EJE.pdb   ---------------------------------------MRIILLGAPGAGKGTQAQFIMEK
## pdbs/1E4Y.pdb   ---------------------------------------MRIILLGALVAGKGTQAQFIMEK
## pdbs/3X2S.pdb   ---------------------------------------MRIILLGAPGAGKGTQAQFIMEK
## pdbs/6HAP.pdb   ---------------------------------------MRIILLGAPGAGKGTQAQFIMEK
## pdbs/6HAM.pdb   ---------------------------------------MRIILLGAPGAGKGTQAQFIMEK
## pdbs/4K46.pdb   ---------------------------------------MRIILLGAPGAGKGTQAQFIMAK
## pdbs/4NP6.pdb   KEAAAGKT--QYLKFDGTKQVSEVSADIAKALA--NAMRIILLGAPGAGKGTQAQFIMEK
## pdbs/3GMT.pdb   ---------------------------------------MRLILLGAPGAGKGTQANFIKEK
## pdbs/4PZL.pdb   NFSSTNTKIPKYIKINGDQAVEKVSQDIFDQLNKRNAMRIILLGAPGAGKGTQAKIIEQK
##                                                       **^*****  *******  *  *
##               421        .         .         .         .         .         480
##
##               481        .         .         .         .         .         540
## pdbs/1AKE.pdb   YGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKERIAQEDCRNGFLL
## pdbs/4X8M.pdb   YGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKERIAQEDCRNGFLL
## pdbs/6S36.pdb   YGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKERIAQEDCRNGFLL
## pdbs/6RZE.pdb   YGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKERIAQEDCRNGFLL
## pdbs/4X8H.pdb   YGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKERIAQEDCRNGFLL
## pdbs/3HPR.pdb   YGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKERIAQEDCRNGFLL
## pdbs/1E4V.pdb   YGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKERIAQEDCRNGFLL
```

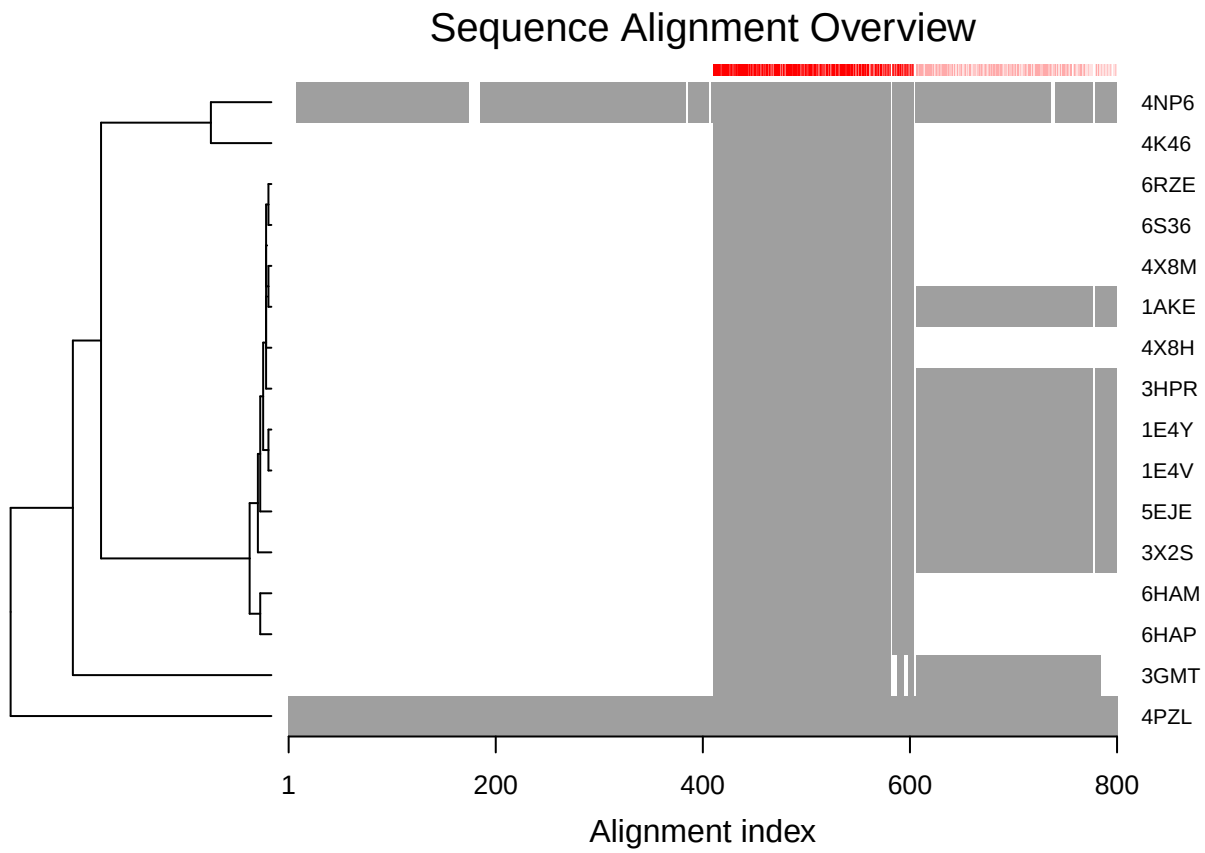
```
## pdbs/5EJE.pdb   YGIPQISTGDMLRAAVKSGSELGKQAKDIMDACKLVTDELVIALVKERIAQEDCRNGFLL
## pdbs/1E4Y.pdb   YGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKERIAQEDCRNGFLL
## pdbs/3X2S.pdb   YGIPQISTGDMLRAAVKSGSELGKQAKDIMDCGKLVTDELVIALVKERIAQEDSRNGFLL
## pdbs/6HAP.pdb   YGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVRERICQEDSRNGFLL
## pdbs/6HAM.pdb   YGIPQISTGDMLRAAIKSGSELGKQAKDIMDAGKLVTDEIIIALVKERICQEDSRNGFLL
## pdbs/4K46.pdb   FGIPQISTGDMLRAAIKAGTELGKQAKSVIDAGQLVSDDIILGLVKERIAQDDCAKGFLL
## pdbs/4NP6.pdb   FGIPQISTGDMLRAAIKAGTELGKQAKAVIDAGQLVSDDIILGLIKERIAQADCEKGFLL
## pdbs/3GMT.pdb   FGIPQISTGDMLRAAVKAGTPLGVEAKTYMDEGKLVPDSLIIGLVKERLKEADCANGYLF
## pdbs/4PZL.pdb   YNIAHISTGDMIRETIKSGSALGQELKKVLDAGELVSDEFIIKIVKDRISKNDCNNGFLL
##                 ^ *  ******^*  ^* *^ **   *  ^*   ** *  ^^ ^^^^*^   *   *^*
##               481        .         .         .         .         .        540
##
##               541        .         .         .         .         .        600
## pdbs/1AKE.pdb   DGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVE
## pdbs/4X8M.pdb   DGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVE
## pdbs/6S36.pdb   DGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDKIVGRRVHAPSGRVYHVKFNPPKVE
## pdbs/6RZE.pdb   DGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDAIVGRRVHAPSGRVYHVKFNPPKVE
## pdbs/4X8H.pdb   DGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVE
## pdbs/3HPR.pdb   DGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVE
## pdbs/1E4V.pdb   DGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVE
## pdbs/5EJE.pdb   DGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVE
## pdbs/1E4Y.pdb   DGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVE
## pdbs/3X2S.pdb   DGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVE
## pdbs/6HAP.pdb   DGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVE
## pdbs/6HAM.pdb   DGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVE
## pdbs/4K46.pdb   DGFPRTIPQADGLKEVGVVVDYVIEFDVADSVIVERMAGRRAHLASGRTYHNVYNPPKVE
## pdbs/4NP6.pdb   DGFPRTIPQADGLKEMGINVDYVIEFDVADDVIVERMAGRRAHLPSGRTYHVVYNPPKVE
## pdbs/3GMT.pdb   DGFPRTIAQADAMKEAGVAIDYVLEIDVPFSEIIERMSGRRTHPASGRTYHVKFNPPKVE
## pdbs/4PZL.pdb   DGVPRTIPQAQELDKLGVNIDYIVEVDVADNLLIERITGRRIHPASGRTYHTKFNPPKVA
##                 ** **** **  ^   *^ ^**^^* **    ^^^ ^ *** *  *** **  ^*****
##               541        .         .         .         .         .        600
##
##               601        .         .         .         .         .        660
## pdbs/1AKE.pdb   GKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKP
## pdbs/4X8M.pdb   GKDDVTGEELTTRKDDQEETVRKRLVEWHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKP
## pdbs/6S36.pdb   GKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKP
## pdbs/6RZE.pdb   GKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKP
## pdbs/4X8H.pdb   GKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAALIGYYSKEAEAGNT--KYAKVDGTKP
## pdbs/3HPR.pdb   GKDDGTGEELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKP
## pdbs/1E4V.pdb   GKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKP
## pdbs/5EJE.pdb   GKDDVTGEELTTRKDDQEECVRKRLVEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKP
## pdbs/1E4Y.pdb   GKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKP
## pdbs/3X2S.pdb   GKDDVTGEELTTRKDDQEETVRKRLCEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKP
## pdbs/6HAP.pdb   GKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKP
## pdbs/6HAM.pdb   GKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKP
## pdbs/4K46.pdb   GKDDVTGEDLVIREDDKEETVLARLGVYHNQTAPLIAYYGKEAEAGNT--QYLKFDGTKA
## pdbs/4NP6.pdb   GKDDVTGEDLVIREDDKEETVRARLNVYHTQTAPLIEYYGKEAAAGKT--QYLKFDGTKQ
## pdbs/3GMT.pdb   GKDDVTGEPLVQRDDDKEETVKKRLDVYEAQTKPLITYYGDWARRGAE-------NGLKA
## pdbs/4PZL.pdb   DKDDVTGEPLITRTDDNEDTVKQRLSVYHAQTAKLIDFYRNFSSTNTKIPKYIKINGDQA
##                  *** *** *  * ** *^ *  **  ^   *  ** ^*               *
##               601        .         .         .         .         .        660
##
##               661        .         .         .         .         .        720
## pdbs/1AKE.pdb   VAEVRADLEKILG--MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSEL
```

```
## pdbs/4X8M.pdb   VAEVRADLEKILG-------------------------------------------------
## pdbs/6S36.pdb   VAEVRADLEKILG-------------------------------------------------
## pdbs/6RZE.pdb   VAEVRADLEKILG-------------------------------------------------
## pdbs/4X8H.pdb   VAEVRADLEKILG-------------------------------------------------
## pdbs/3HPR.pdb   VAEVRADLEKILG--MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSEL
## pdbs/1E4V.pdb   VAEVRADLEKILG--MRIILLGAPVAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSEL
## pdbs/5EJE.pdb   VAEVRADLEKILG--MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSEL
## pdbs/1E4Y.pdb   VAEVRADLEKILG--MRIILLGALVAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSEL
## pdbs/3X2S.pdb   VAEVRADLEKILG--MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSEL
## pdbs/6HAP.pdb   VCEVRADLEKILG-------------------------------------------------
## pdbs/6HAM.pdb   VCEVRADLEKILG-------------------------------------------------
## pdbs/4K46.pdb   VAEVSAELEKALA-------------------------------------------------
## pdbs/4NP6.pdb   VSEVSADIAKALA-AMRIILLGAPGAGKGTQAQFIMEKFGIPQISTGDMLRAAIKAGTEL
## pdbs/3GMT.pdb   PA-----YRKISG--MRLILLGAPGAGKGTQANFIKEKFGIPQISTGDMLRAAVKAGTPL
## pdbs/4PZL.pdb   VEKVSQDIFDQLNKAMRIILLGAPGAGKGTQAKIIEQKYNIAHISTGDMIRETIKSGSAL
##
##               661       .         .         .         .         .         720
##
##               721       .         .         .         .         .         780
## pdbs/1AKE.pdb   GKQAKDIMDAGKLVTDELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDY
## pdbs/4X8M.pdb   ------------------------------------------------------------
## pdbs/6S36.pdb   ------------------------------------------------------------
## pdbs/6RZE.pdb   ------------------------------------------------------------
## pdbs/4X8H.pdb   ------------------------------------------------------------
## pdbs/3HPR.pdb   GKQAKDIMDAGKLVTDELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDY
## pdbs/1E4V.pdb   GKQAKDIMDAGKLVTDELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDY
## pdbs/5EJE.pdb   GKQAKDIMDACKLVTDELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDY
## pdbs/1E4Y.pdb   GKQAKDIMDAGKLVTDELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDY
## pdbs/3X2S.pdb   GKQAKDIMDCGKLVTDELVIALVKERIAQEDSRNGFLLDGFPRTIPQADAMKEAGINVDY
## pdbs/6HAP.pdb   ------------------------------------------------------------
## pdbs/6HAM.pdb   ------------------------------------------------------------
## pdbs/4K46.pdb   ------------------------------------------------------------
## pdbs/4NP6.pdb   GKQAKAVIDAGQLVSDDIILGLIKERIAQADCEKGFLLDGFPRTIPQADGLKEMGINVDY
## pdbs/3GMT.pdb   GVEAKTYMDEGKLVPDSLIIGLVKERLKEADCANGYLFDGFPRTIAQADAMKEAGVAIDY
## pdbs/4PZL.pdb   GQELKKVLDAGELVSDEFIIKIVKDRISKNDCNNGFLLDGVPRTIPQAQELDKLGVNIDY
##
##               721       .         .         .         .         .         780
##
##               781       .         .         .         .         .         840
## pdbs/1AKE.pdb   VLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVR
## pdbs/4X8M.pdb   ------------------------------------------------------------
## pdbs/6S36.pdb   ------------------------------------------------------------
## pdbs/6RZE.pdb   ------------------------------------------------------------
## pdbs/4X8H.pdb   ------------------------------------------------------------
## pdbs/3HPR.pdb   VLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDGTGEELTTRKDDQEETVR
## pdbs/1E4V.pdb   VLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVR
## pdbs/5EJE.pdb   VLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEECVR
## pdbs/1E4Y.pdb   VLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVR
## pdbs/3X2S.pdb   VLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVR
## pdbs/6HAP.pdb   ------------------------------------------------------------
## pdbs/6HAM.pdb   ------------------------------------------------------------
## pdbs/4K46.pdb   ------------------------------------------------------------
## pdbs/4NP6.pdb   VIEFDVADDVIVERMAGRRAHLPSGRTYHVVYNPPKVEGKD----EDLVIREDDKEETVR
## pdbs/3GMT.pdb   VLEIDVPFSEIIERMSGRRTHPASGRTYHVKFNPPKVEGKDDVTGEPLVQRDDDKEETVK
```

```
## pdbs/4PZL.pdb   IVEVDVADNLLIERITGRRIHPASGRTYHTKFNPPKVADKDDVTGEPLITRTDDNEDTVK
##
##               781        .         .         .         .         .         840
##
##               841        .         .         .         .         . 892
## pdbs/1AKE.pdb   KRLVEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKPVAEVRADLEKILG-
## pdbs/4X8M.pdb   ----------------------------------------------------
## pdbs/6S36.pdb   ----------------------------------------------------
## pdbs/6RZE.pdb   ----------------------------------------------------
## pdbs/4X8H.pdb   ----------------------------------------------------
## pdbs/3HPR.pdb   KRLVEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKPVAEVRADLEKILG-
## pdbs/1E4V.pdb   KRLVEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKPVAEVRADLEKILG-
## pdbs/5EJE.pdb   KRLVEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKPVAEVRADLEKILG-
## pdbs/1E4Y.pdb   KRLVEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKPVAEVRADLEKILG-
## pdbs/3X2S.pdb   KRLCEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTKPVAEVRADLEKILG-
## pdbs/6HAP.pdb   ----------------------------------------------------
## pdbs/6HAM.pdb   ----------------------------------------------------
## pdbs/4K46.pdb   ----------------------------------------------------
## pdbs/4NP6.pdb   ARLNVYHTQTAPLIEYYGKEAAAGKT--QYLKFDGTKQVSEVSADIAKALA-
## pdbs/3GMT.pdb   KRLDVYEAQTKPLITYYGDWARRGAKAPAYRKIS------------------
## pdbs/4PZL.pdb   QRLSVYHAQTAKLIDFYRNFSSTNTKIPKYIKINGDQAVEKVSQDIFDQLNK
##
##               841        .         .         .         .         . 892
##
## Call:
##   pdbaln(files = files, fit = TRUE, exefile = "msa")
##
## Class:
##   pdbs, fasta
##
## Alignment dimensions:
##   16 sequence rows; 892 position columns (204 non-gap, 688 gap)
##
## + attr: xyz, resno, b, chain, id, ali, resid, sse, call
```

```
# Vector containing PDB codes for figure axis
ids <- basename.pdb(pdbs$id)

# Draw schematic alignment
plot(pdbs, labels=ids)
```



Collect annotation for each entry:

```
anno <- pdb.annotate(ids)
unique(anno$source)
```

```
## [1] "Escherichia coli"
## [2] "Escherichia coli K-12"
## [3] "Escherichia coli O139:H28 str. E24377A"
## [4] "Escherichia coli str. K-12 substr. MDS42"
## [5] "Photobacterium profundum"
## [6] "Vibrio cholerae O1 biovar El Tor str. N16961"
## [7] "Burkholderia pseudomallei 1710b"
## [8] "Francisella tularensis subsp. tularensis SCHU S4"
```

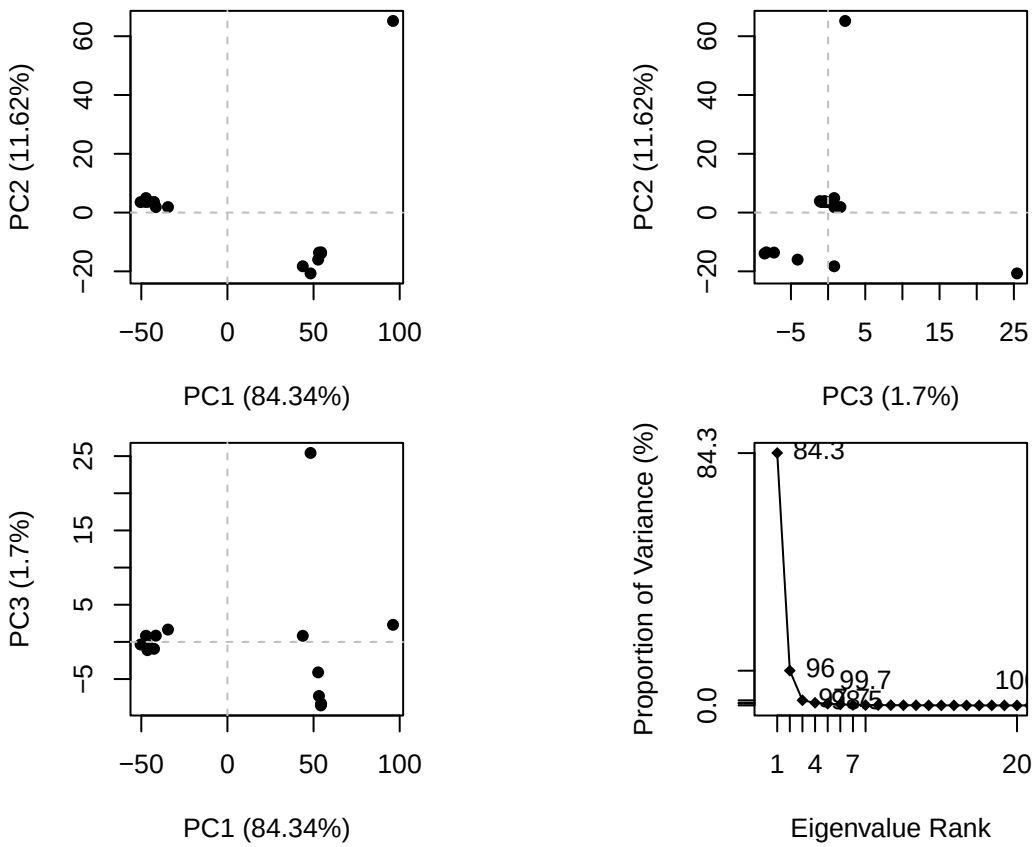
```
head(anno)
```

```
##        structureId chainId macromoleculeType chainLength experimentalTechnique
## 1AKE_A        1AKE       A           Protein         214                X-ray
## 1AKE_B        1AKE       B           Protein         214                X-ray
## 4X8M_A        4X8M       A           Protein         214                X-ray
## 6S36_A        6S36       A           Protein         214                X-ray
## 6RZE_A        6RZE       A           Protein         214                X-ray
## 4X8H_A        4X8H       A           Protein         214                X-ray
##        resolution       scopDomain                   pfam         ligandId
## 1AKE_A       2.00 Adenylate kinase Adenylate kinase (ADK)              AP5
```

```
## 1AKE_B       2.00 Adenylate kinase Adenylate kinase (ADK)              AP5
## 4X8M_A       2.60             <NA> Adenylate kinase (ADK)             <NA>
## 6S36_A       1.60             <NA> Adenylate kinase (ADK) CL (3),NA,MG (2)
## 6RZE_A       1.69             <NA> Adenylate kinase (ADK)    NA (3),CL (2)
## 4X8H_A       2.50             <NA> Adenylate kinase (ADK)             <NA>
##                                        ligandName           source
## 1AKE_A              BIS(ADENOSINE)-5'-PENTAPHOSPHATE Escherichia coli
## 1AKE_B              BIS(ADENOSINE)-5'-PENTAPHOSPHATE Escherichia coli
## 4X8M_A                                          <NA> Escherichia coli
## 6S36_A CHLORIDE ION (3),SODIUM ION,MAGNESIUM ION (2) Escherichia coli
## 6RZE_A               SODIUM ION (3),CHLORIDE ION (2) Escherichia coli
## 4X8H_A                                          <NA> Escherichia coli
##
## 1AKE_A STRUCTURE OF THE COMPLEX BETWEEN ADENYLATE KINASE FROM ESCHERICHIA COLI AND THE INHIBITOR AP5A
## 1AKE_B STRUCTURE OF THE COMPLEX BETWEEN ADENYLATE KINASE FROM ESCHERICHIA COLI AND THE INHIBITOR AP5A
## 4X8M_A
## 6S36_A
## 6RZE_A
## 4X8H_A
##                                     citation rObserved  rFree  rWork
## 1AKE_A  Muller, C.W., et al. J Mol Biol (1992)    0.1960     NA 0.1960
## 1AKE_B  Muller, C.W., et al. J Mol Biol (1992)    0.1960     NA 0.1960
## 4X8M_A Kovermann, M., et al. Nat Commun (2015)    0.2491 0.3089 0.2463
## 6S36_A   Rogne, P., et al. Biochemistry (2019)    0.1632 0.2356 0.1594
## 6RZE_A   Rogne, P., et al. Biochemistry (2019)    0.1865 0.2350 0.1819
## 4X8H_A Kovermann, M., et al. Nat Commun (2015)    0.1961 0.2895 0.1914
##        spaceGroup
## 1AKE_A  P 21 2 21
## 1AKE_B  P 21 2 21
## 4X8M_A    C 1 2 1
## 6S36_A    C 1 2 1
## 6RZE_A    C 1 2 1
## 4X8H_A    C 1 2 1
```

Time for PCA. We will use no tthe `prcomp()` function from base R, but the `pca()` function from the `bio3d` package as this one is designed to work nicely with biomolecular data.

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



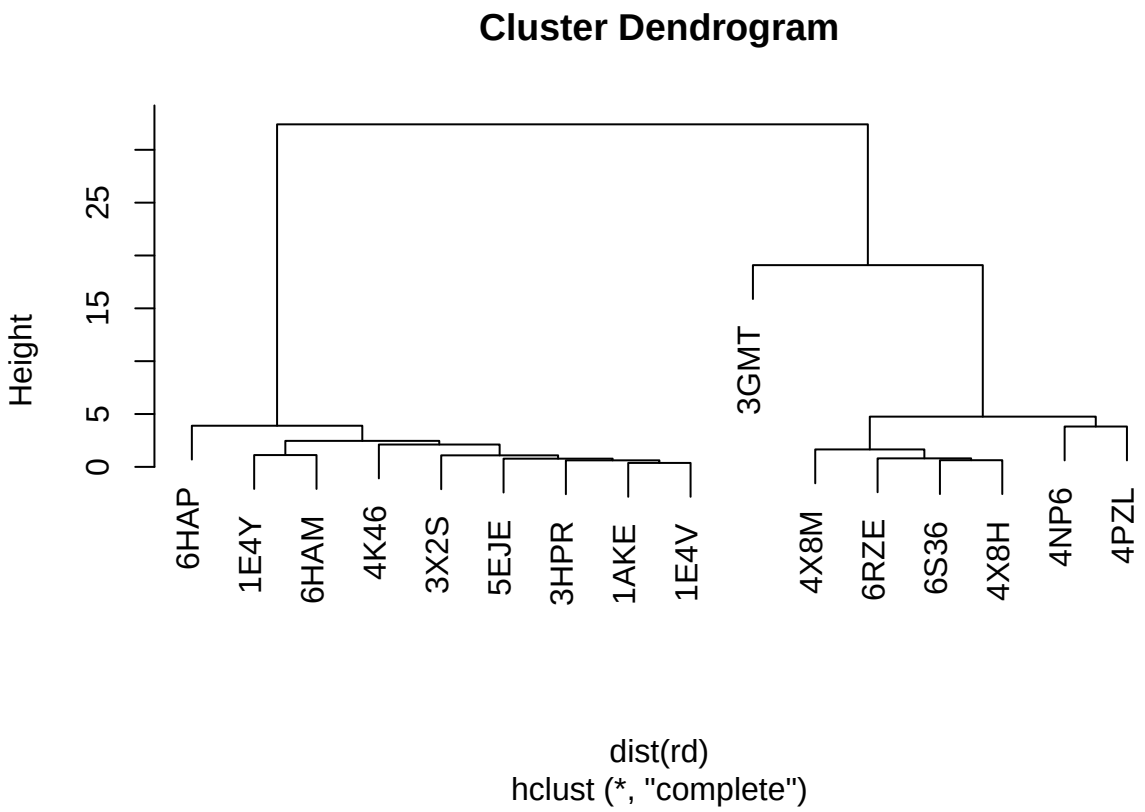
We can now focus in on PC1 vs PC2

```
# Calculate RMSD
rd <- rmsd(pdbs)
```

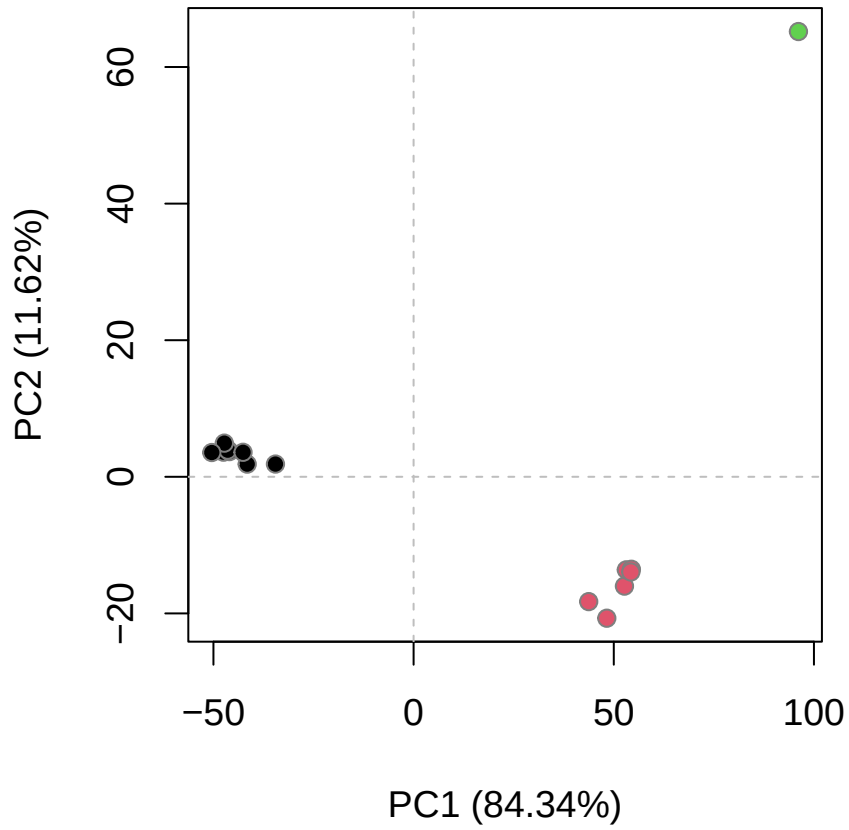
```
## Warning in rmsd(pdbs): No indices provided, using the 204 non NA positions
```

```
# Structure-based clustering
hc.rd <- hclust(dist(rd))
grps.rd <- cutree(hc.rd, k=3)

plot(hc.rd)
```




```
plot(pc.xray, 1:2, col="grey50", bg=grps.rd, pch=21, cex=1)
```



Further Visualization

To visualize the major structural variations in the ensemble the function `mktrj()` can be used to generate a trajectory PDB file by interpolating along a give PC (eigenvector):

```
# Visualize first principal component
mktrj(pc.xray, pc=1, file="pc_1.pdb")
```

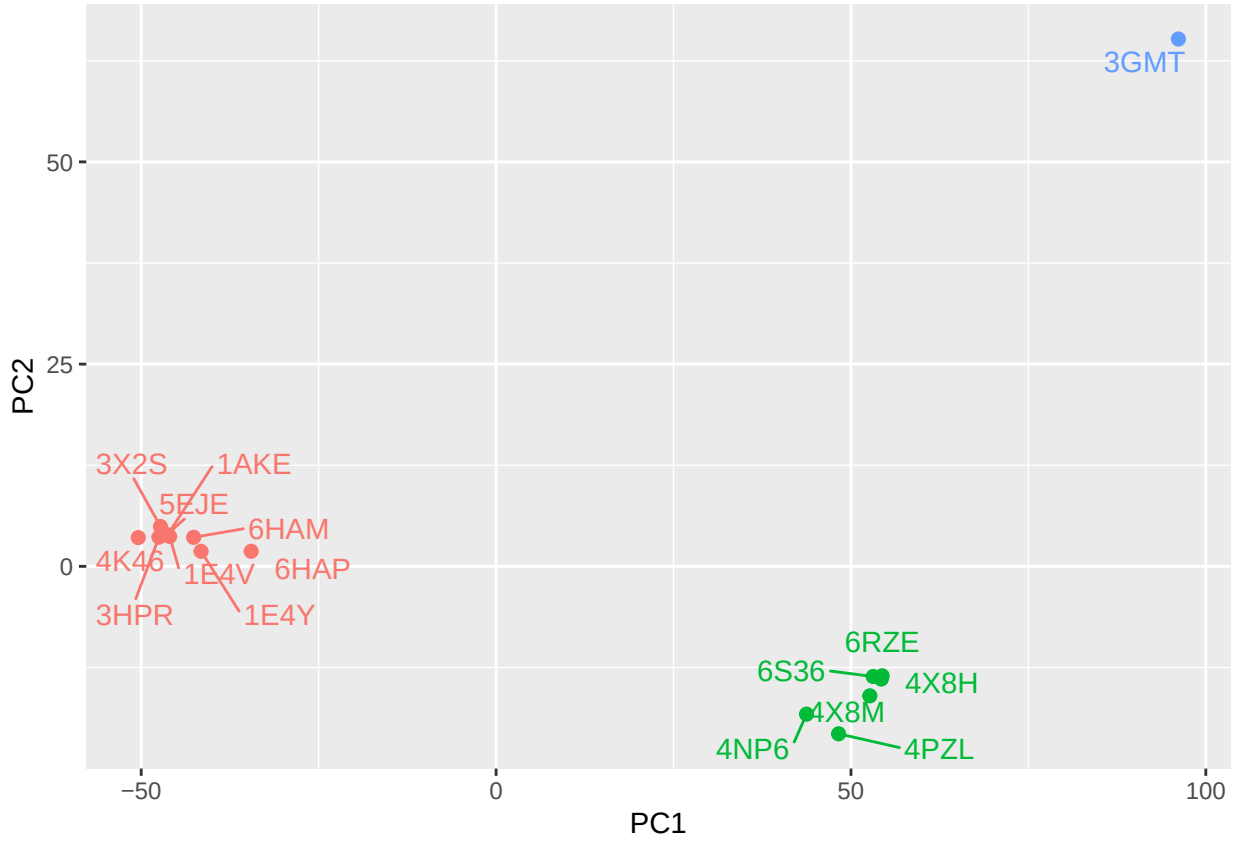
We can now open this trajectory file in Molstar to view a wee movie of the major differences (i.e. displacements) in the structure set as we move along PC1.

```
#Plotting results with ggplot2
library(ggplot2)
library(ggrepel)

df <- data.frame(PC1=pc.xray$z[,1],
                 PC2=pc.xray$z[,2],
                 col=as.factor(grps.rd),
                 ids=ids)

p <- ggplot(df) +
  aes(PC1, PC2, col=col, label=ids) +
  geom_point(size=2) +
  geom_text_repel(max.overlaps = 20) +
```

```
  theme(legend.position = "none")
p
```



Normal Mode Analysis

Function `nma()` provides normal mode analysis (NMA) on both single structures (if given a singe PDB input object) or the complete structure ensemble (if provided with a PDBS input object). This facilitates characterizing and comparing flexibility profiles of related protein structures.