

Class12: Structural Bioinformatics II

Rui Huang (PID: A15606522)

11/4/2021

#Comparative Analysis of protein structures using the bio3d

```
library(bio3d)
```

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

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

```
pdb
```

```
##
```

```
## Call: read.pdb(file = "1hel")
```

```
##
```

```
## Total Models#: 1
```

```
## Total Atoms#: 1186, XYZs#: 3558 Chains#: 1 (values: A)
```

```
##
```

```
## Protein Atoms#: 1001 (residues/Calpha atoms#: 129)
```

```
## Nucleic acid Atoms#: 0 (residues/phosphate atoms#: 0)
```

```
##
```

```
## Non-protein/nucleic Atoms#: 185 (residues: 185)
```

```
## Non-protein/nucleic resid values: [ HOH (185) ]
```

```
##
```

```
## Protein sequence:
```

```
## KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGDYDYGILQINS
```

```
## RWWCNDGRTPGSRNLCNIPCSALLSSDITASVNC AKKIVSDGNMNAWVAWRNRCKGTDV
```

```
## QAWIRGCRL
```

```
##
```

```
## + attr: atom, xyz, seqres, helix, sheet,
```

```
## calpha, remark, call
```

```
#Q7: there are 129 aa residues
```

```
#Q8:water
```

```
#Q9:1
```

```
#Q10:"msa" is only found on BioConductor
```

```
#Q11:"Grantlab/bio3d-view" is not on CRAN or BioConductor
```

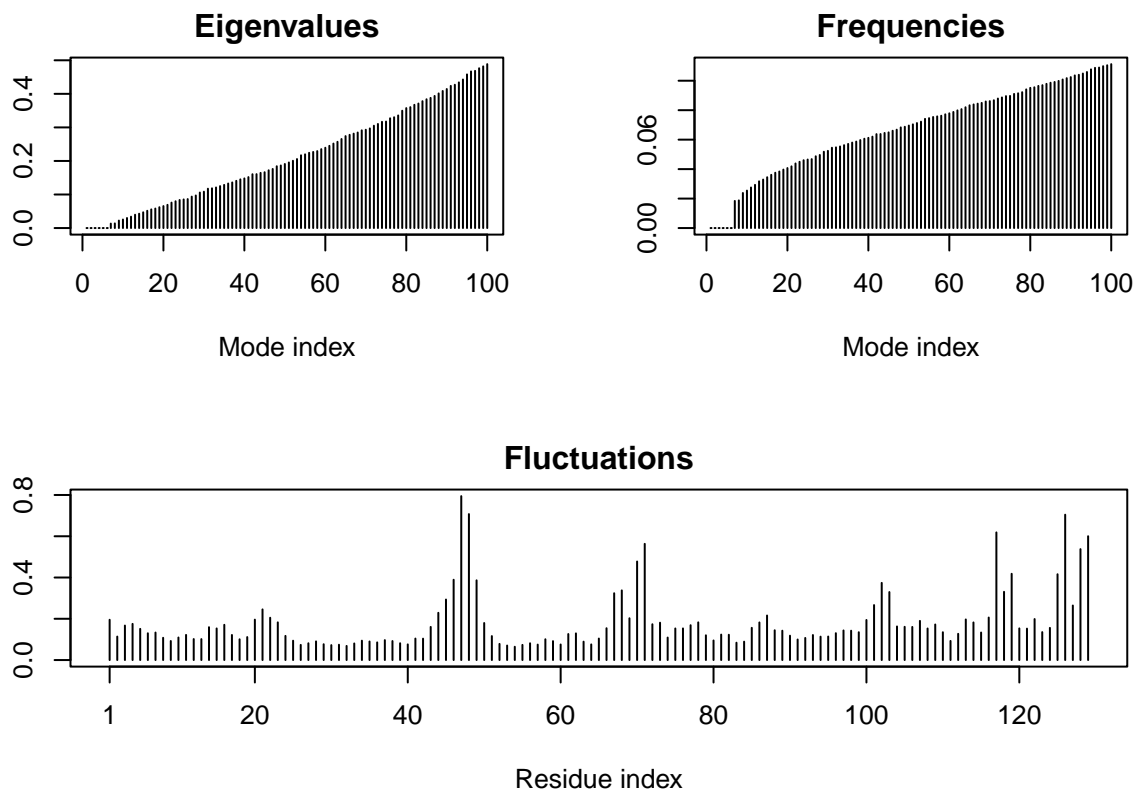
```
#Q12:True
```

Normal Mode Analysis to predict the dynamics (flexibility) of the enzyme.

```
modes <- nma(pdb)
```

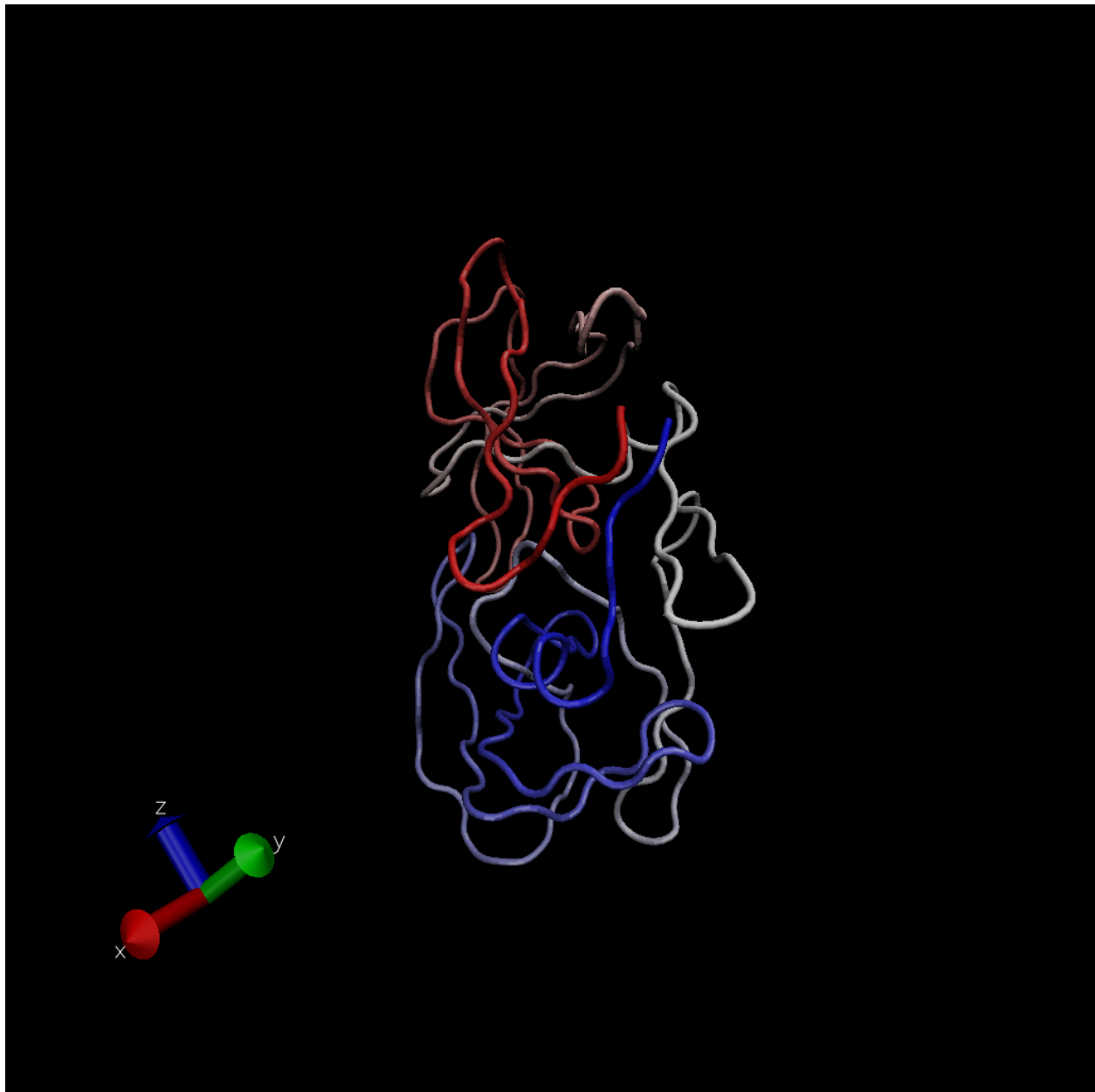
```
## Building Hessian... Done in 0.044 seconds.  
## Diagonalizing Hessian... Done in 0.172 seconds.
```

```
plot(modes)
```



Make a “move” of its prediction motion. call this “trajectory”

```
mktrj(modes, file="nma.pdb")
```



```
#Analysis of ADK
```

```
aa <- get.seq("lake_A")
```

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

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

```
aa
```

```
##          1          .          .          .          .          .          60
## pdb|1AKE|A  MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLV
##          1          .          .          .          .          .          60
```

```
##
##          61          .          .          .          .          120
## pdb|1AKE|A  DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRI
##          61          .          .          .          .          120
##
##          121         .          .          .          .          180
## pdb|1AKE|A  VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTRKDDQEETVRKRLVEYHQMTAPLIG
##          121         .          .          .          .          180
##
##          181         .          .          .          214
## pdb|1AKE|A  YYSKEAEAGNTKYAKVDGTPVAEVRADLEKILG
##          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
```

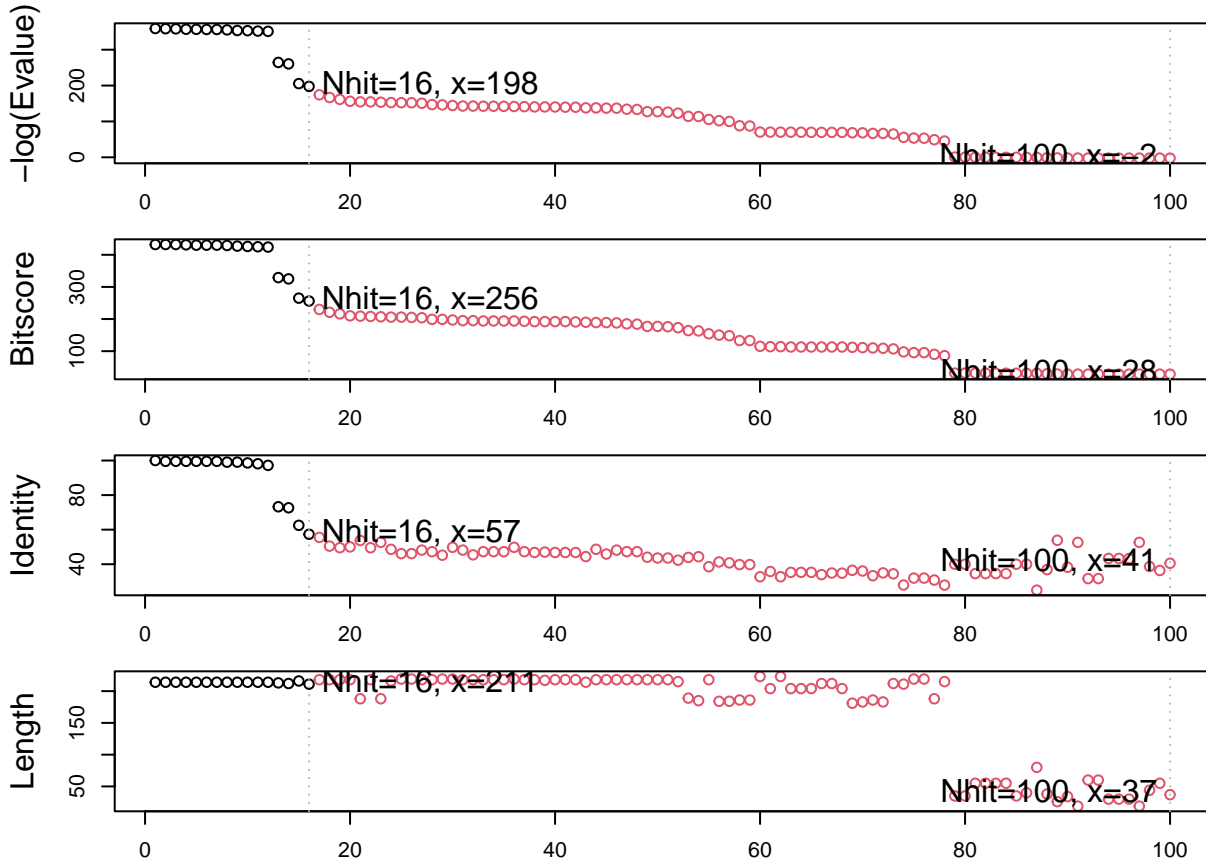
#Q13: there are 214 amino acids.

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

```
## Searching ... please wait (updates every 5 seconds) RID = SHNPF39R013
## .....
## Reporting 100 hits
```

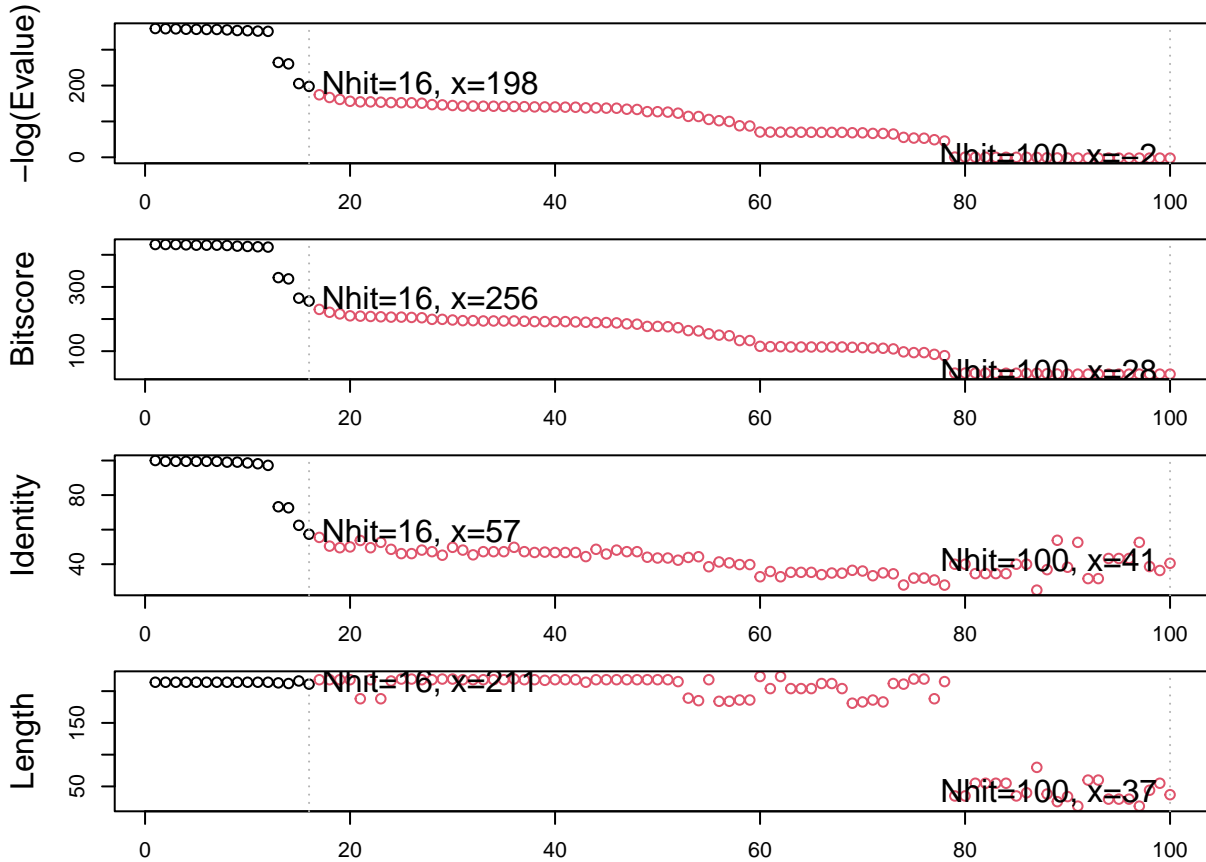
```
plot(blast)
```

```
## * Possible cutoff values: 197 -3
##           Yielding Nhits: 16 100
##
## * Chosen cutoff value of: 197
##           Yielding Nhits: 16
```



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

```
## * Possible cutoff values: 197 -3
##      Yielding Nhits: 16 100
##
## * Chosen cutoff value of: 197
##      Yielding Nhits: 16
```



```
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"
```

```
hits <- NULL
```

```
hits$pdb.id <- c('1AKE_A','4X8M_A','6S36_A','6RZE_A','4X8H_A','3HPR_A','1E4V_A','5EJE_A','1E4Y_A','3X2S_A')
```

```
files <- get.pdb(hits$pdb.id, path="pdb", splot="T", gzip="T")
```

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

```
pdbs <- pdbaln(files, fit=T)
```

```
## Reading PDB files:
## pdb/1AKE.pdb
## pdb/4X8M.pdb
## pdb/6S36.pdb
## pdb/6RZE.pdb
## pdb/4X8H.pdb
## pdb/3HPR.pdb
## pdb/1E4V.pdb
## pdb/5EJE.pdb
## pdb/1E4Y.pdb
## pdb/3X2S.pdb
## pdb/6HAP.pdb
## pdb/6HAM.pdb
## pdb/4K46.pdb
## pdb/4NP6.pdb
## pdb/3GMT.pdb
```

```

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

```

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

```

## Reading PDB files:
## pdb/1AKE.pdb
## pdb/4X8M.pdb
## pdb/6S36.pdb
## pdb/6RZE.pdb
## pdb/4X8H.pdb
## pdb/3HPR.pdb
## pdb/1E4V.pdb
## pdb/5EJE.pdb
## pdb/1E4Y.pdb
## pdb/3X2S.pdb
## pdb/6HAP.pdb

```



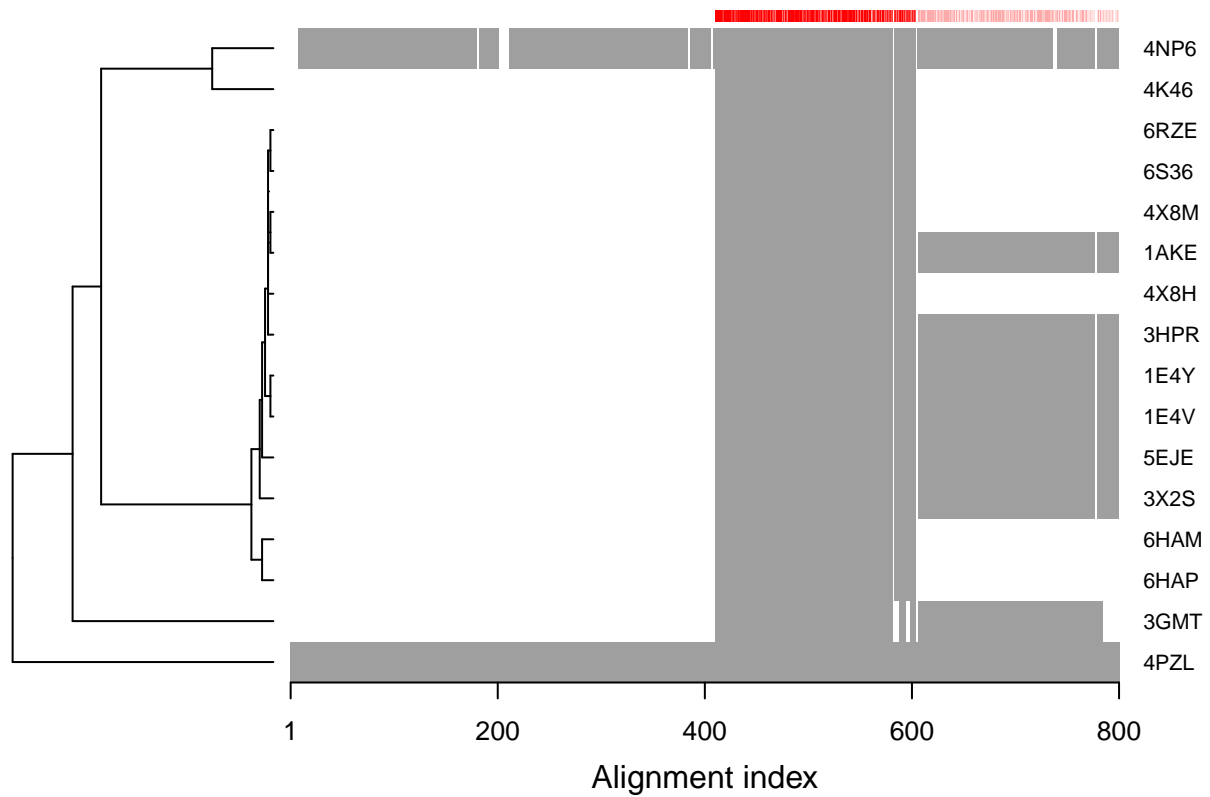
```

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

ids <- basename.pdb(pdb$id)
plot(pdb, labels=ids)

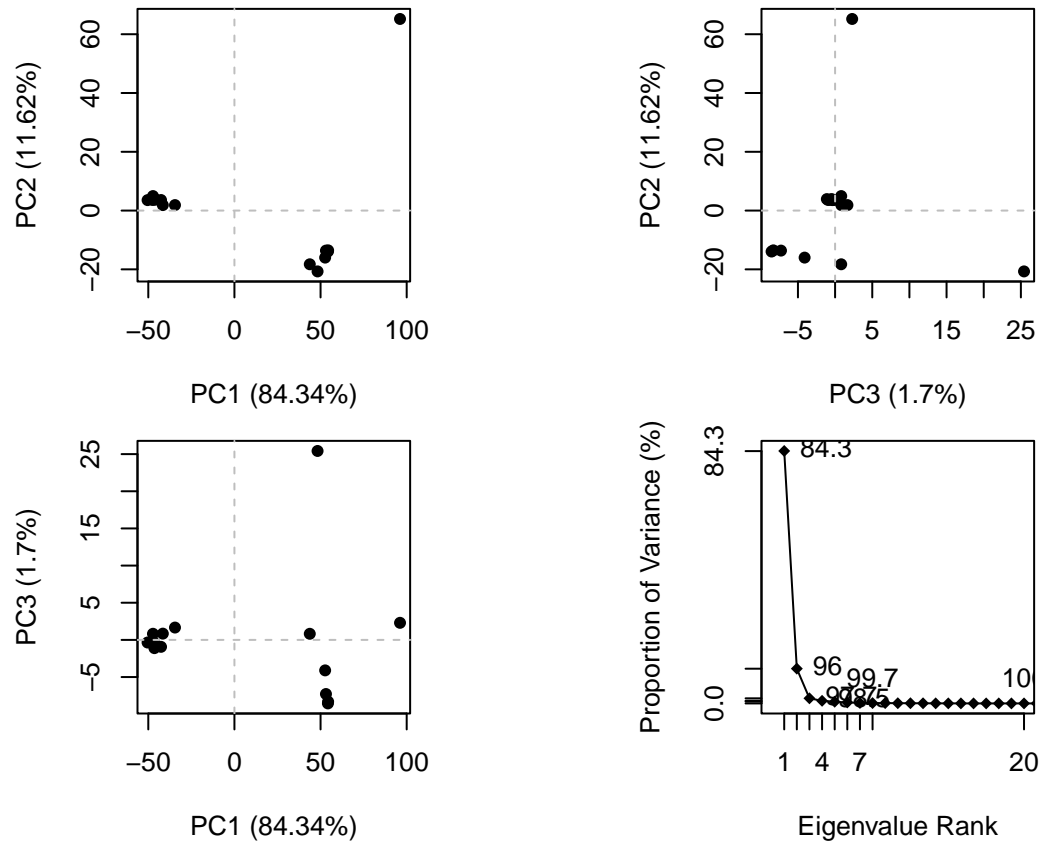
```

Sequence Alignment Overview



#PCA

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



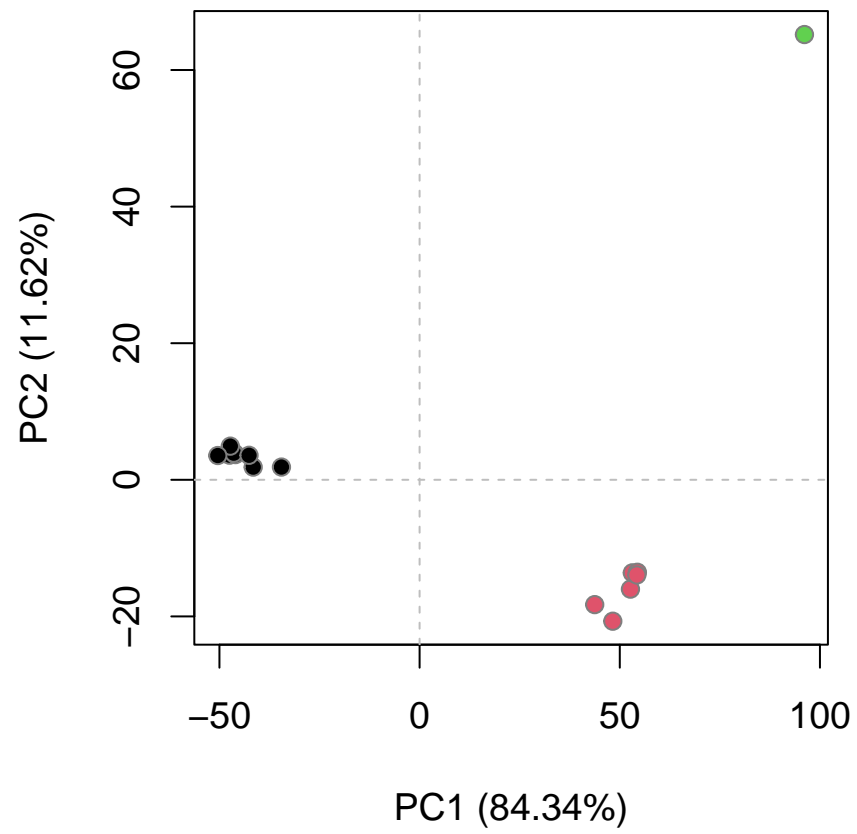
```
rd <- rmsd(pdb)
```

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

```
hc.rd <- hclust(dist(rd))
```

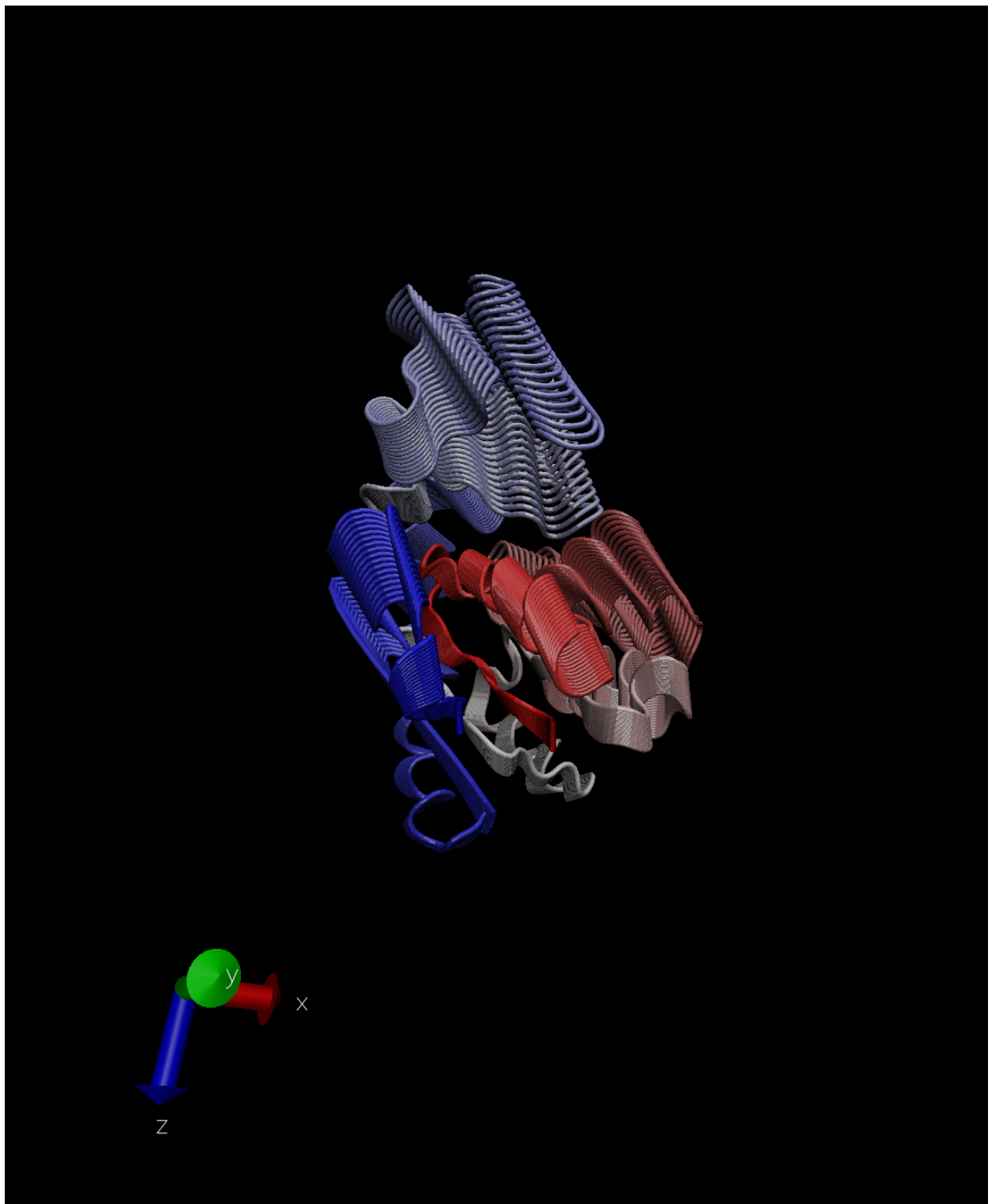
```
grps.rd <- cutree(hc.rd, k=3)
```

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



Make a trajectory visualization of the motion captured by the PC1

```
pc1 <- mktrj(pc.xray,pc=1,file="pc_1.pdb")
```



Normal mode analysis

```
modes <- nma(pdb)
```

```
## Warning in nma.pdb(pdb): 1AKE.pdb, 3HPR.pdb, 1E4V.pdb, 5EJE.pdb, 1E4Y.pdb, 3X2S.pdb, 4NP6.pdb, 3GM  
## Fluctuations at neighboring positions may be affected.
```

```
##
## Details of Scheduled Calculation:
## ... 16 input structures
## ... storing 606 eigenvectors for each structure
## ... dimension of x$U.subspace: ( 612x606x16 )
## ... coordinate superposition prior to NM calculation
## ... aligned eigenvectors (gap containing positions removed)
## ... estimated memory usage of final 'eNMA' object: 45.4 Mb
##
## |
```

```
plot(modes,pdb, col=grps.rd)
```

```
## Extracting SSE from pbs$sse attribute
```

