Class12: structual biology II

Tianru Zhang (PID: A15432834)

11/4/2021

```
library(bio3d)
#1hel or 1hsg
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
```

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

198

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

MK1

Q9: How many protein chains are in this structure?

Let's use Bioinformatics methods called NMA(normal model analysis) to predict the dynamics(flexibility) of this enzyme
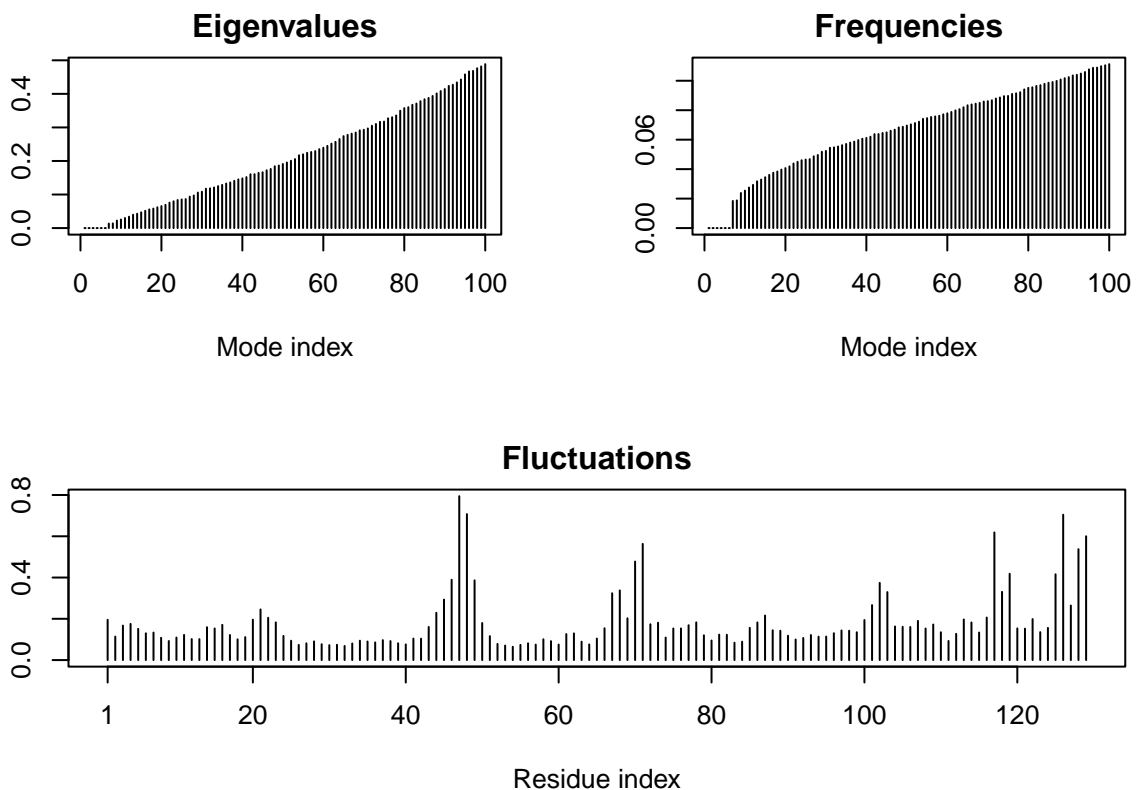
```
pdb<-read.pdb("1hel")
```

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

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

```
## Building Hessian...      Done in 0.04 seconds.
## Diagonalizing Hessian...     Done in 0.34 seconds.
```
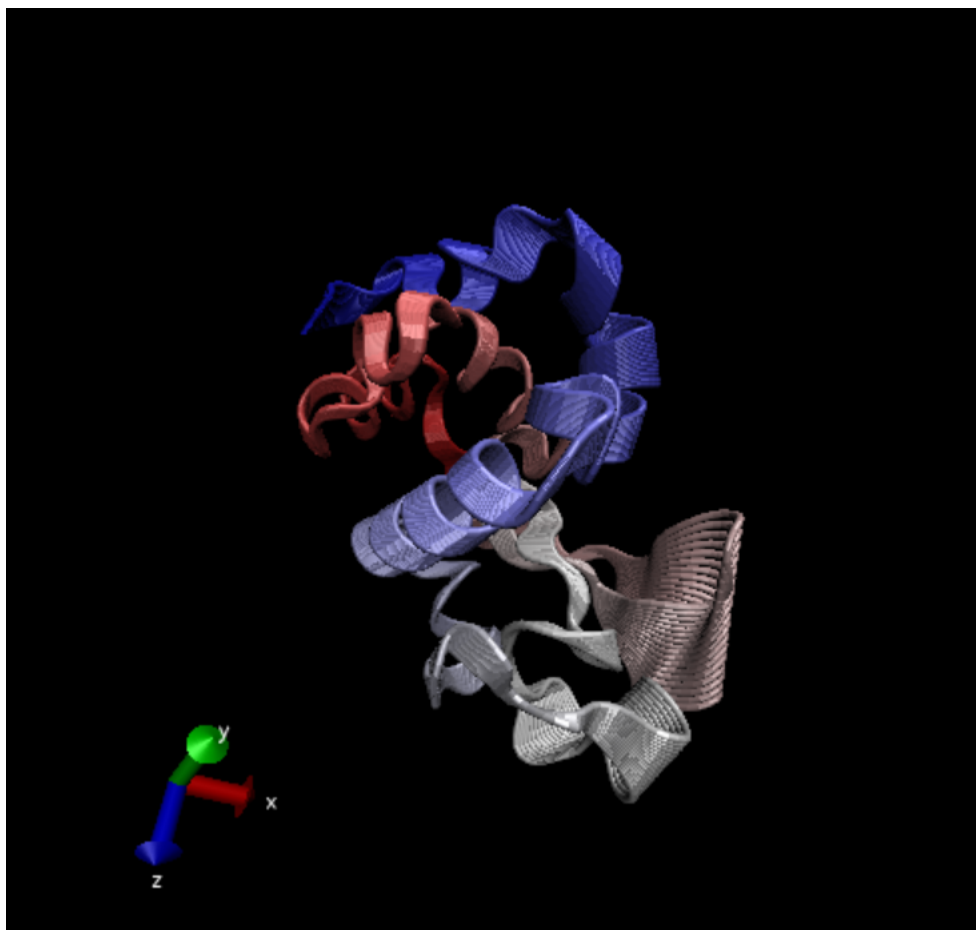
```
plot(modes)
```

Eigenvalues

Frequencies

Fluctuations

Make a move of the predicted motion called "trajectory"

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

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

msa

Q11. Which of the 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

#Analysis of ADK

```
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

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

```
#hits<-plot(blast)
#hits$pdb.id
```

```
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_

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/
## 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/
## 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/
## 4K46.pdb exists. Skipping download

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

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

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

##   |                                                                       |
```

##Align and superpose structures

```r
#install.packages("ggrepel")
#install.packages("devtools")
#install.packages("BiocManager")

#BiocManager::install("msa")
#devtools::install_bitbucket("Grantlab/bio3d-view")
pdbs <- pdbaln(files, fit = TRUE, exefile="msa")
```

```
## Reading PDB files:
## pdbs/split_chain/1AKE_A.pdb
## pdbs/split_chain/4X8M_A.pdb
## pdbs/split_chain/6S36_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/4K46_A.pdb
## pdbs/split_chain/4NP6_A.pdb
## pdbs/split_chain/3GMT_A.pdb
## pdbs/split_chain/4PZL_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
## ....
##
## 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/4X8M_A.pdb
## pdb/seq: 3   name: pdbs/split_chain/6S36_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 4   name: pdbs/split_chain/6RZE_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 5   name: pdbs/split_chain/4X8H_A.pdb
## pdb/seq: 6   name: pdbs/split_chain/3HPR_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 7   name: pdbs/split_chain/1E4V_A.pdb
## pdb/seq: 8   name: pdbs/split_chain/5EJE_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 9   name: pdbs/split_chain/1E4Y_A.pdb
## pdb/seq: 10   name: pdbs/split_chain/3X2S_A.pdb
## pdb/seq: 11   name: pdbs/split_chain/6HAP_A.pdb
## pdb/seq: 12   name: pdbs/split_chain/6HAM_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 13   name: pdbs/split_chain/4K46_A.pdb
##    PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 14   name: pdbs/split_chain/4NP6_A.pdb
## pdb/seq: 15   name: pdbs/split_chain/3GMT_A.pdb
## pdb/seq: 16   name: pdbs/split_chain/4PZL_A.pdb
```

```
# Vector containing PDB codes for figure axis
pdbs
```

```
##                                          1         .         .         .        40
## [Truncated_Name:1]1AKE_A.pdb            ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
## [Truncated_Name:2]4X8M_A.pdb            ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
## [Truncated_Name:3]6S36_A.pdb            ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
## [Truncated_Name:4]6RZE_A.pdb            ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
## [Truncated_Name:5]4X8H_A.pdb            ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
## [Truncated_Name:6]3HPR_A.pdb            ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
## [Truncated_Name:7]1E4V_A.pdb            ----------MRIILLGAPVAGKGTQAQFIMEKYGIPQIS
## [Truncated_Name:8]5EJE_A.pdb            ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
## [Truncated_Name:9]1E4Y_A.pdb            ----------MRIILLGALVAGKGTQAQFIMEKYGIPQIS
## [Truncated_Name:10]3X2S_A.pdb           ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
## [Truncated_Name:11]6HAP_A.pdb           ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
## [Truncated_Name:12]6HAM_A.pdb           ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
## [Truncated_Name:13]4K46_A.pdb           ----------MRIILLGAPGAGKGTQAQFIMAKFGIPQIS
## [Truncated_Name:14]4NP6_A.pdb           --------NAMRIILLGAPGAGKGTQAQFIMEKFGIPQIS
## [Truncated_Name:15]3GMT_A.pdb           ----------MRLILLGAPGAGKGTQANFIKEKFGIPQIS
## [Truncated_Name:16]4PZL_A.pdb           TENLYFQSNAMRIILLGAPGAGKGTQAKIIEQKYNIAHIS
##                                                   **^*****  *******  *   *^ *   **
##                                          1         .         .         .        40
##
##                                          41        .         .         .        80
## [Truncated_Name:1]1AKE_A.pdb             TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
## [Truncated_Name:2]4X8M_A.pdb             TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
## [Truncated_Name:3]6S36_A.pdb             TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
## [Truncated_Name:4]6RZE_A.pdb             TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
## [Truncated_Name:5]4X8H_A.pdb             TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
## [Truncated_Name:6]3HPR_A.pdb             TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
## [Truncated_Name:7]1E4V_A.pdb             TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
## [Truncated_Name:8]5EJE_A.pdb             TGDMLRAAVKSGSELGKQAKDIMDACKLVTDELVIALVKE
## [Truncated_Name:9]1E4Y_A.pdb             TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
## [Truncated_Name:10]3X2S_A.pdb            TGDMLRAAVKSGSELGKQAKDIMDCGKLVTDELVIALVKE
## [Truncated_Name:11]6HAP_A.pdb            TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVRE
## [Truncated_Name:12]6HAM_A.pdb            TGDMLRAAIKSGSELGKQAKDIMDAGKLVTDEIIIALVKE
## [Truncated_Name:13]4K46_A.pdb            TGDMLRAAIKAGTELGKQAKSVIDAGQLVSDDIILGLVKE
## [Truncated_Name:14]4NP6_A.pdb            TGDMLRAAIKAGTELGKQAKAVIDAGQLVSDDIILGLIKE
## [Truncated_Name:15]3GMT_A.pdb            TGDMLRAAVKAGTPLGVEAKTYMDEGKLVPDSLIIGLVKE
## [Truncated_Name:16]4PZL_A.pdb            TGDMIRETIKSGSALGQELKKVLDAGELVSDEFIIKIVKD
##                                          ****^*  ^* *^ **    *  ^*   ** *  ^^ ^^^^
##                                          41        .         .         .        80
##
##                                          81        .         .         .       120
## [Truncated_Name:1]1AKE_A.pdb             RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
## [Truncated_Name:2]4X8M_A.pdb             RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
## [Truncated_Name:3]6S36_A.pdb             RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
## [Truncated_Name:4]6RZE_A.pdb             RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
## [Truncated_Name:5]4X8H_A.pdb             RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
## [Truncated_Name:6]3HPR_A.pdb             RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
## [Truncated_Name:7]1E4V_A.pdb             RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
## [Truncated_Name:8]5EJE_A.pdb             RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
## [Truncated_Name:9]1E4Y_A.pdb             RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
## [Truncated_Name:10]3X2S_A.pdb            RIAQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
## [Truncated_Name:11]6HAP_A.pdb            RICQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
## [Truncated_Name:12]6HAM_A.pdb            RICQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
## [Truncated_Name:13]4K46_A.pdb            RIAQDDCAKGFLLDGFPRTIPQADGLKEVGVVVDYVIEFD
```

```
## [Truncated_Name:14]4NP6_A.pdb    RIAQADCEKGFLLDGFPRTIPQADGLKEMGINVDYVIEFD
## [Truncated_Name:15]3GMT_A.pdb    RLKEADCANGYLFDGFPRTIAQADAMKEAGVAIDYVLEID
## [Truncated_Name:16]4PZL_A.pdb    RISKNDCNNGFLLDGVPRTIPQAQELDKLGVNIDYIVEVD
##                                  *^    *   *^* ** **** **  ^   *^ ^**^^* *
##                                  81         .         .         .        120
##
## 
##                                  121        .         .         .        160
## [Truncated_Name:1]1AKE_A.pdb     VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
## [Truncated_Name:2]4X8M_A.pdb     VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
## [Truncated_Name:3]6S36_A.pdb     VPDELIVDKIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
## [Truncated_Name:4]6RZE_A.pdb     VPDELIVDAIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
## [Truncated_Name:5]4X8H_A.pdb     VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
## [Truncated_Name:6]3HPR_A.pdb     VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDGTG
## [Truncated_Name:7]1E4V_A.pdb     VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
## [Truncated_Name:8]5EJE_A.pdb     VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
## [Truncated_Name:9]1E4Y_A.pdb     VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
## [Truncated_Name:10]3X2S_A.pdb    VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
## [Truncated_Name:11]6HAP_A.pdb    VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
## [Truncated_Name:12]6HAM_A.pdb    VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
## [Truncated_Name:13]4K46_A.pdb    VADSVIVERMAGRRAHLASGRTYHNVYNPPKVEGKDDVTG
## [Truncated_Name:14]4NP6_A.pdb    VADDVIVERMAGRRAHLPSGRTYHVVYNPPKVEGKDDVTG
## [Truncated_Name:15]3GMT_A.pdb    VPFSEIIERMSGRRTHPASGRTYHVKFNPPKVEGKDDVTG
## [Truncated_Name:16]4PZL_A.pdb    VADNLLIERITGRRIHPASGRTYHTKFNPPKVADKDDVTG
##                                  *   ^^^ ^ *** *  *** **  ^*****  *** **
##                                  121        .         .         .        160
##
## 
##                                  161        .         .         .        200
## [Truncated_Name:1]1AKE_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
## [Truncated_Name:2]4X8M_A.pdb     EELTTRKDDQEETVRKRLVEWHQMTAPLIGYYSKEAEAGN
## [Truncated_Name:3]6S36_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
## [Truncated_Name:4]6RZE_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
## [Truncated_Name:5]4X8H_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAALIGYYSKEAEAGN
## [Truncated_Name:6]3HPR_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
## [Truncated_Name:7]1E4V_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
## [Truncated_Name:8]5EJE_A.pdb     EELTTRKDDQEECVRKRLVEYHQMTAPLIGYYSKEAEAGN
## [Truncated_Name:9]1E4Y_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
## [Truncated_Name:10]3X2S_A.pdb    EELTTRKDDQEETVRKRLCEYHQMTAPLIGYYSKEAEAGN
## [Truncated_Name:11]6HAP_A.pdb    EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
## [Truncated_Name:12]6HAM_A.pdb    EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
## [Truncated_Name:13]4K46_A.pdb    EDLVIREDDKEETVLARLGVYHNQTAPLIAYYGKEAEAGN
## [Truncated_Name:14]4NP6_A.pdb    EDLVIREDDKEETVRARLNVYHTQTAPLIEYYGKEAAAGK
## [Truncated_Name:15]3GMT_A.pdb    EPLVQRDDDKEETVKKRLDVYEAQTKPLITYYGDWARRGA
## [Truncated_Name:16]4PZL_A.pdb    EPLITRTDDNEDTVKQRLSVYHAQTAKLIDFYRNFSSTNT
##                                  * *  * ** *^ *  **  ^   *  ** ^*
##                                  161        .         .         .        200
##
## 
##                                  201        .         .   227
## [Truncated_Name:1]1AKE_A.pdb     T--KYAKVDGTKPVAEVRADLEKILG-
## [Truncated_Name:2]4X8M_A.pdb     T--KYAKVDGTKPVAEVRADLEKILG-
## [Truncated_Name:3]6S36_A.pdb     T--KYAKVDGTKPVAEVRADLEKILG-
## [Truncated_Name:4]6RZE_A.pdb     T--KYAKVDGTKPVAEVRADLEKILG-
## [Truncated_Name:5]4X8H_A.pdb     T--KYAKVDGTKPVAEVRADLEKILG-
## [Truncated_Name:6]3HPR_A.pdb     T--KYAKVDGTKPVAEVRADLEKILG-
## [Truncated_Name:7]1E4V_A.pdb     T--KYAKVDGTKPVAEVRADLEKILG-
```

```
## [Truncated_Name:8]5EJE_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
## [Truncated_Name:9]1E4Y_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
## [Truncated_Name:10]3X2S_A.pdb   T--KYAKVDGTKPVAEVRADLEKILG-
## [Truncated_Name:11]6HAP_A.pdb   T--KYAKVDGTKPVCEVRADLEKILG-
## [Truncated_Name:12]6HAM_A.pdb   T--KYAKVDGTKPVCEVRADLEKILG-
## [Truncated_Name:13]4K46_A.pdb   T--QYLKFDGTKAVAEVSAELEKALA-
## [Truncated_Name:14]4NP6_A.pdb   T--QYLKFDGTKQVSEVSADIAKALA-
## [Truncated_Name:15]3GMT_A.pdb   E-------NGLKAPA-----YRKISG-
## [Truncated_Name:16]4PZL_A.pdb   KIPKYIKINGDQAVEKVSQDIFDQLNK
##                                              *
##                            201        .         .         227
##
## Call:
##   pdbaln(files = files, fit = TRUE, exefile = "msa")
##
## Class:
##   pdbs, fasta
##
## Alignment dimensions:
##   16 sequence rows; 227 position columns (204 non-gap, 23 gap)
##
## + attr: xyz, resno, b, chain, id, ali, resid, sse, call
```

```r
ids <- basename.pdb(pdbs$id)

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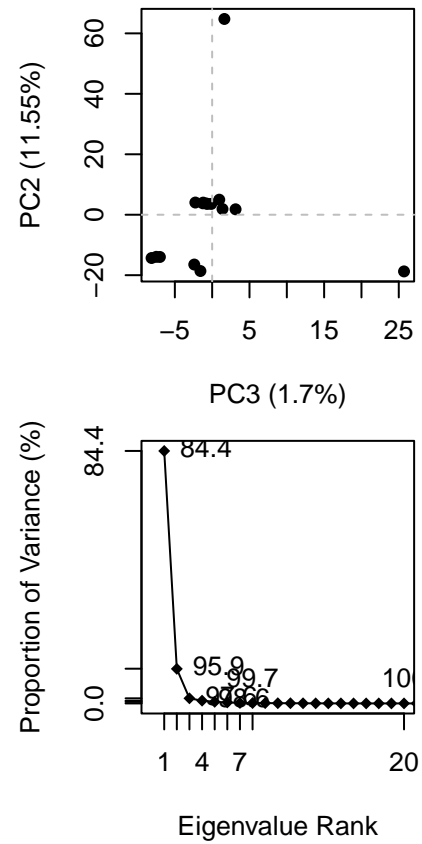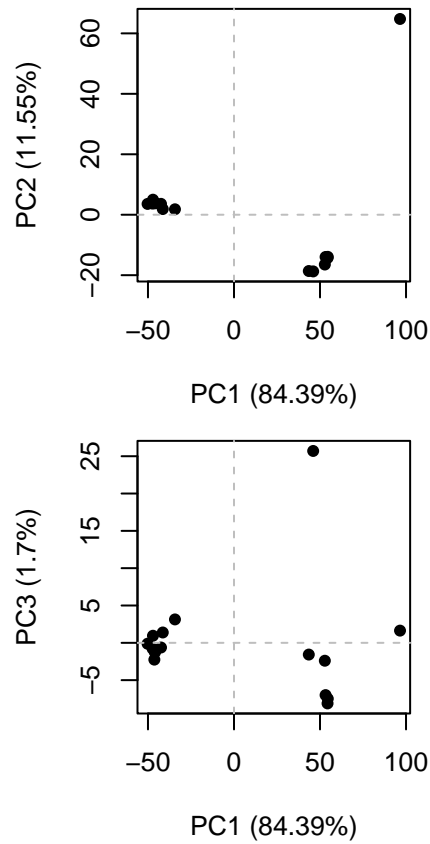
Sequence Alignment Overview
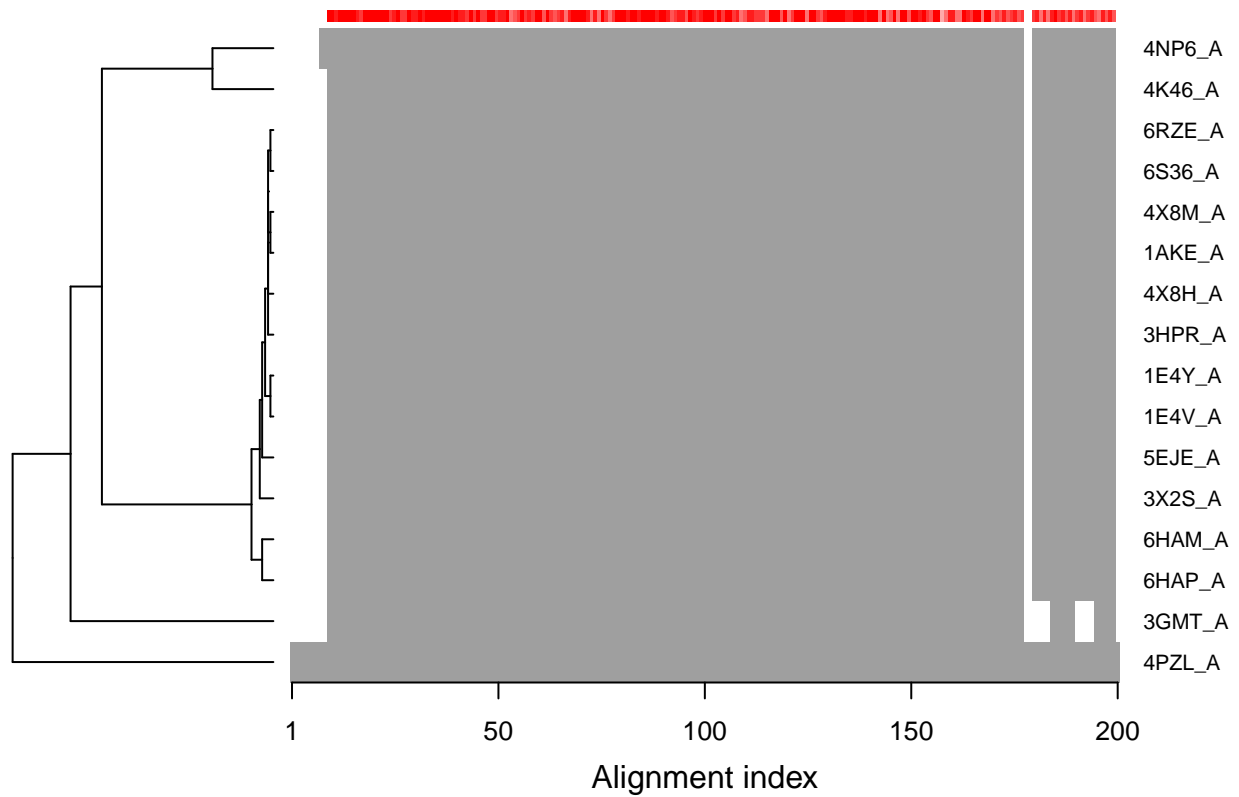

```
library(bio3d.view)
#install.packages("rgl")
library(rgl)

view.pdbs(pdbs)
```

##PCA analysis
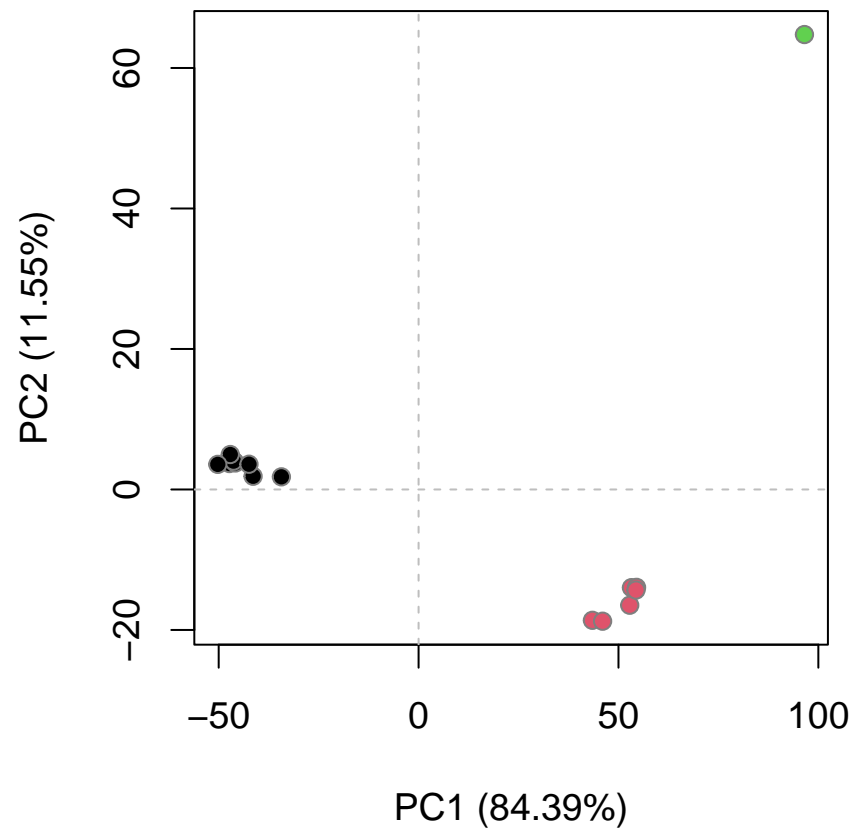
```
# Perform PCA
pc.xray <- pca(pdbs)
plot(pc.xray)
```

PC2 (11.55%)

PC1 (84.39%)

PC2 (11.55%)

PC3 (1.7%)

PC3 (1.7%)

PC1 (84.39%)

Proportion of Variance (%)

84.4

95.9 99.7 10

Eigenvalue Rank

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

```
## Warning in rmsd(pdbs): No indices provided, using the 204 non NA positions
```
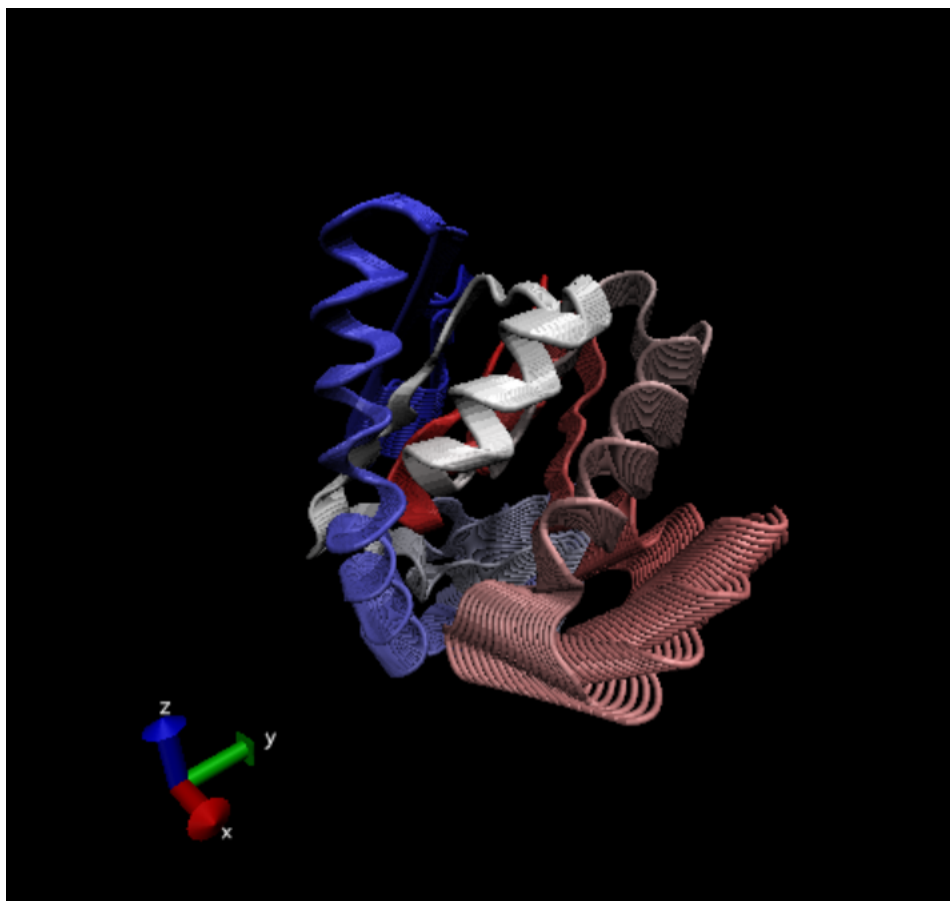
```r
# Structure-based clustering
hc.rd <- hclust(dist(rd))
grps.rd <- cutree(hc.rd, k=3)

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

#Further visualization

```r
# Visualize first principal component
pc1 <- mktrj(pc.xray, pc=1, file="pc_1.pdb")
```

```
view.xyz(pc1)
```

```
## Potential all C-alpha atom structure(s) detected: Using calpha.connectivity()
```

```
view.xyz(pc1, col=vec2color( rmsf(pc1) ))
```

```
## Potential all C-alpha atom structure(s) detected: Using calpha.connectivity()
```
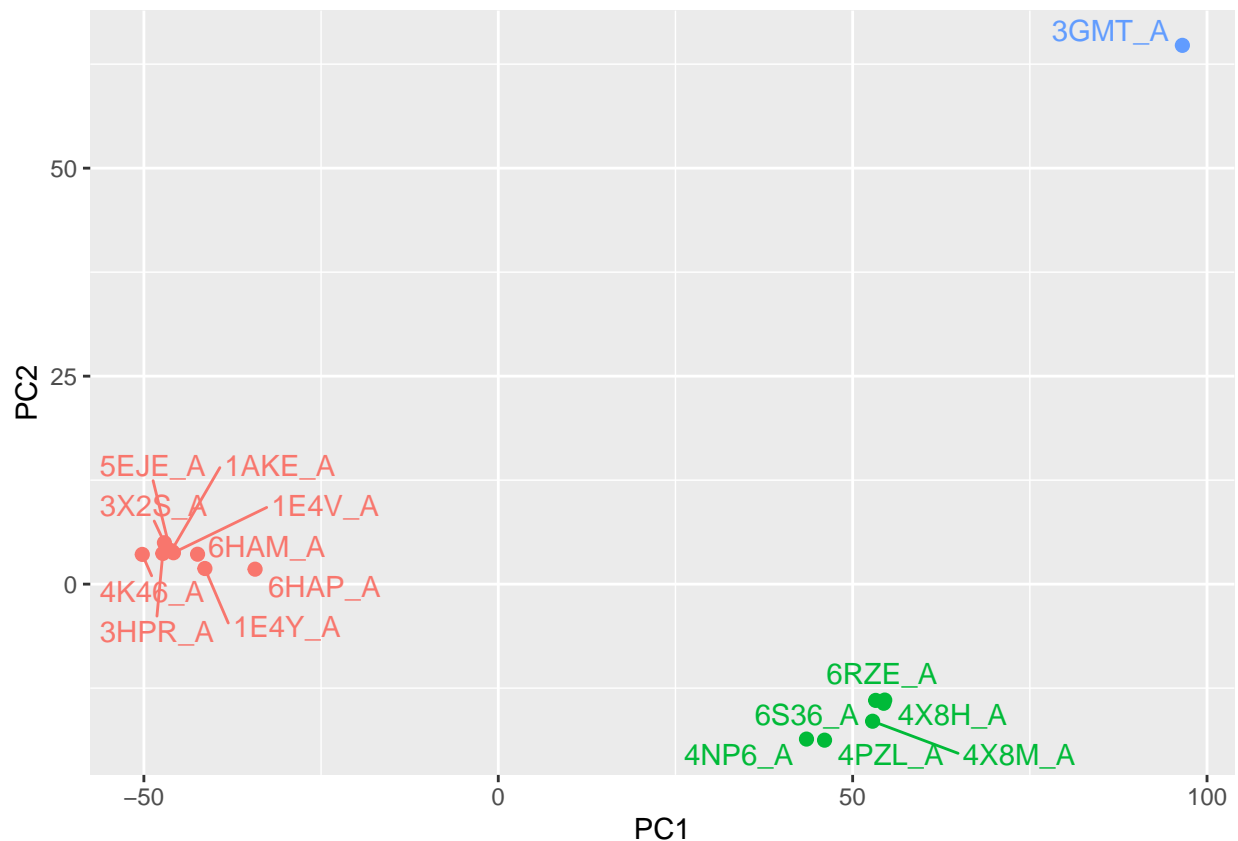
```
#Plotting results with ggplot2
library(ggplot2)
library(ggrepel)

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

```
# NMA of all structures
modes <- nma(pdbs)
```

```
##
## 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
##
##    |                                                        |
```
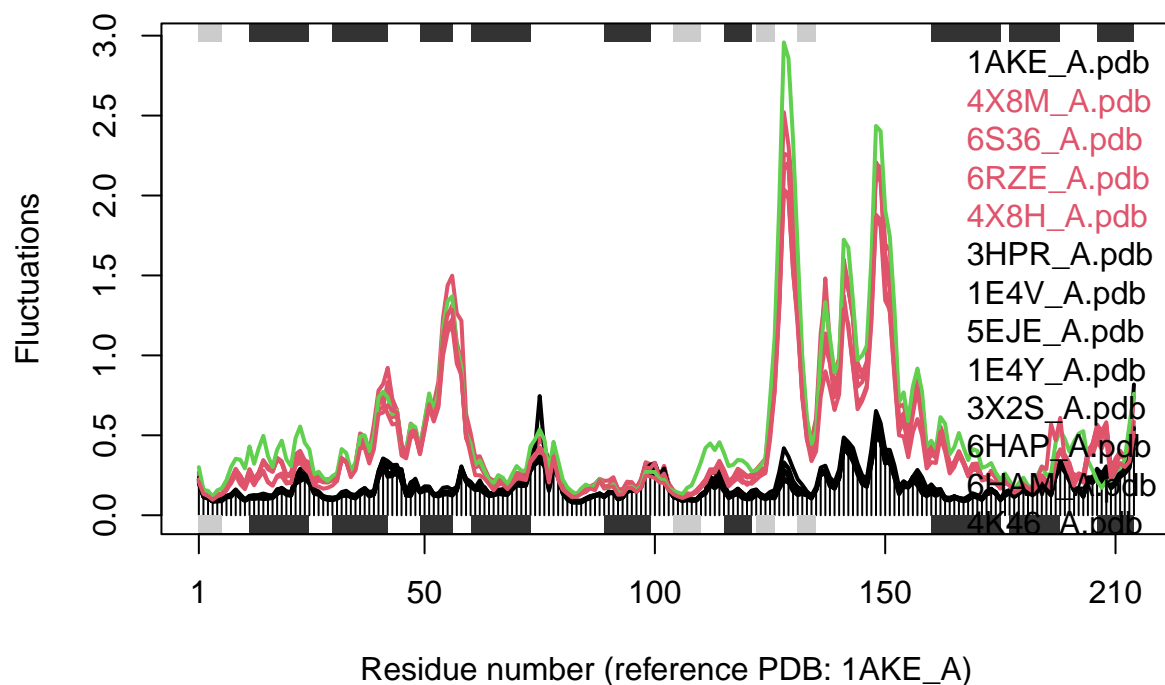
```
plot(modes, pdbs, col=grps.rd)
```

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

Residue number (reference PDB: 1AKE_A)

>Q14. What do you note about this plot? Are the black and colored lines similar or different? Where do you think they differ most and why?

The black and colored lines are different, that the colored lines tend to fluctuate more and the differ most at residue number 50 and 150. Those are the hinge portions of the ligand binding parts of the protein, which have the most specificity and differences.
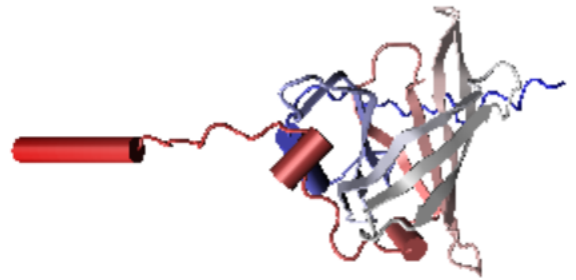
The one below is the png for the "Find a gene project"