class12: Structural Bioinformatics II

Anel A15426506

11/4/2021

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

198

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

HOH >Q9: How many protein chains are in this structure?

2

#Comparative analysis of protein structures

using the bio3d package

```
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:
##       KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINS
##       RWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAWRNRCKGTDV
##       QAWIRGCRL
##
## + attr: atom, xyz, seqres, helix, sheet,
##         calpha, remark, call
```

let's use a bioinformarics method called NMA (Normal Mode ANalysis) to predict the dynamics (flexibilty) of this enzyme.
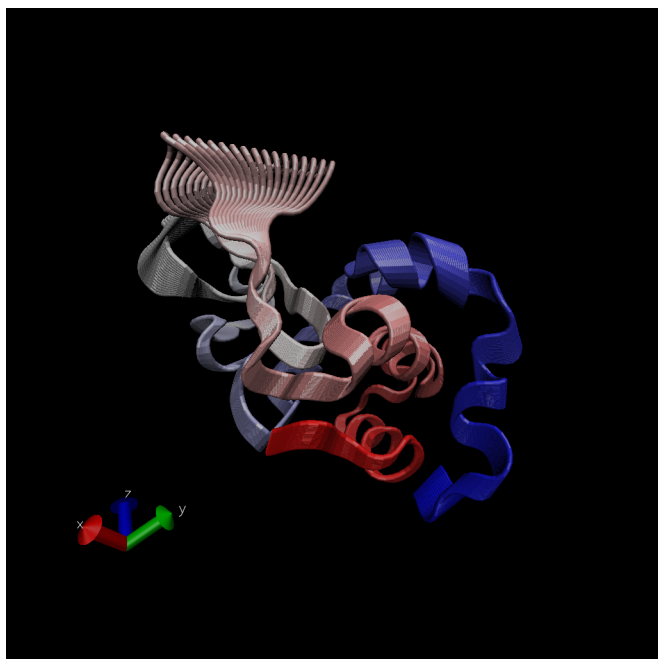
```
modes <- nma(pdb)
```

```
##  Building Hessian...     Done in 0.02 seconds.
##  Diagonalizing Hessian...    Done in 0.22 seconds.
```

```
plot(modes)
```

Eigenvalues

Frequencies

Fluctuations
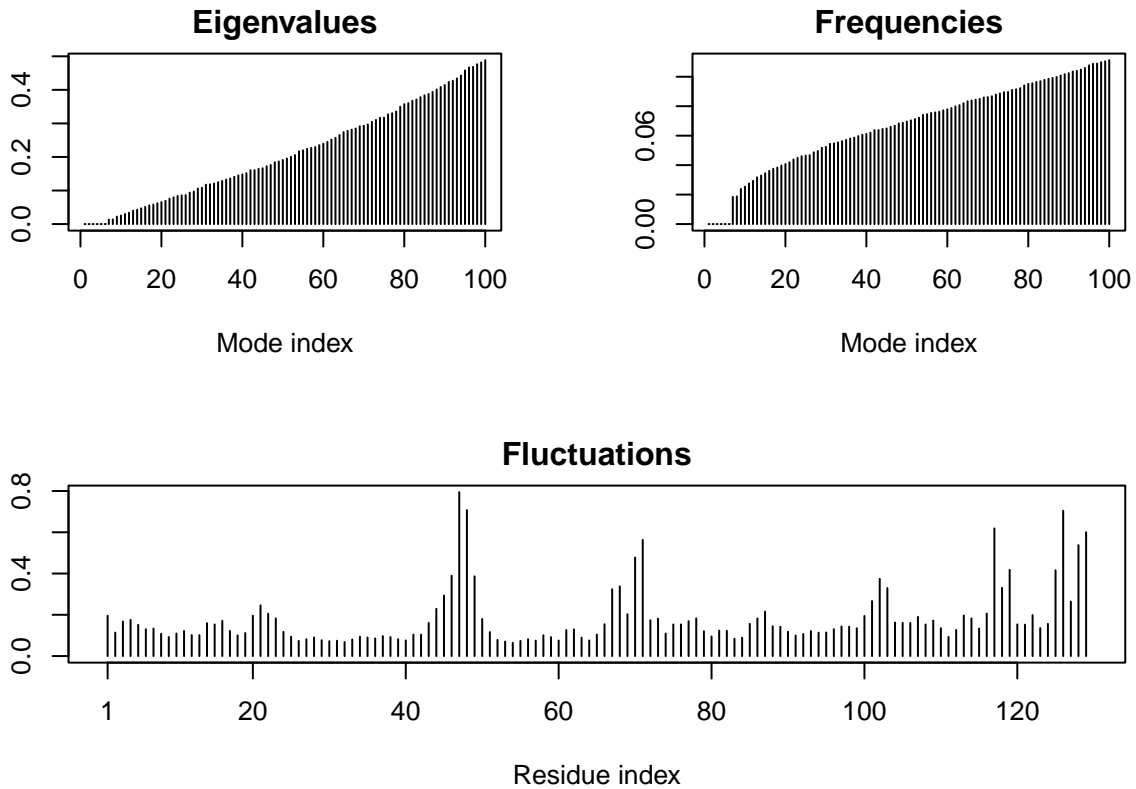
make a "move" of its predicted motion. we often call this a "trajectory"

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

```r
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_
```

download PDB files

```r
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.gz exists. Skipping download

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

##     |                                                                  |
```

Multiple structure alignment

```
pdb <- pdbaln(files, fit= TRUE)
```

```
## 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
```

pdb

```
##                                  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
##                                            **^***** ******* *  *^ *  **
```

5

```
##                                 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)
##
## 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
```
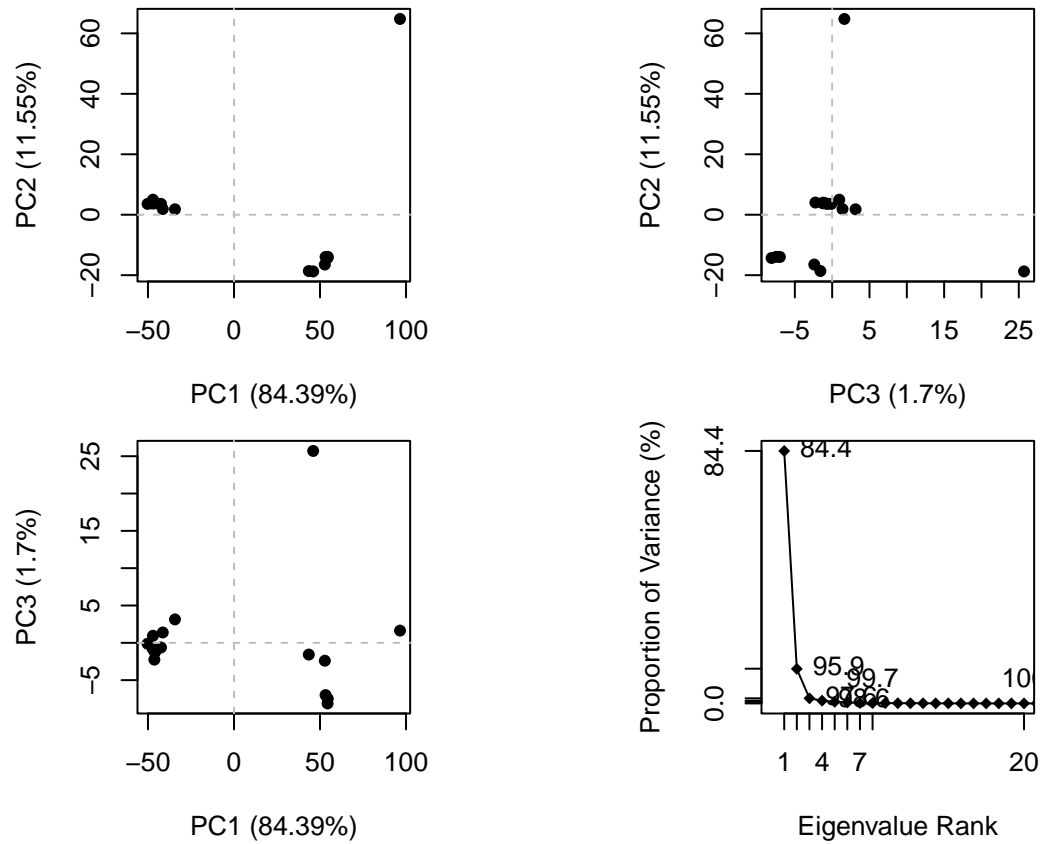
```
ids <- basename.pdb(pdb$id)
```

```
plot(pdb, labels=ids)
```
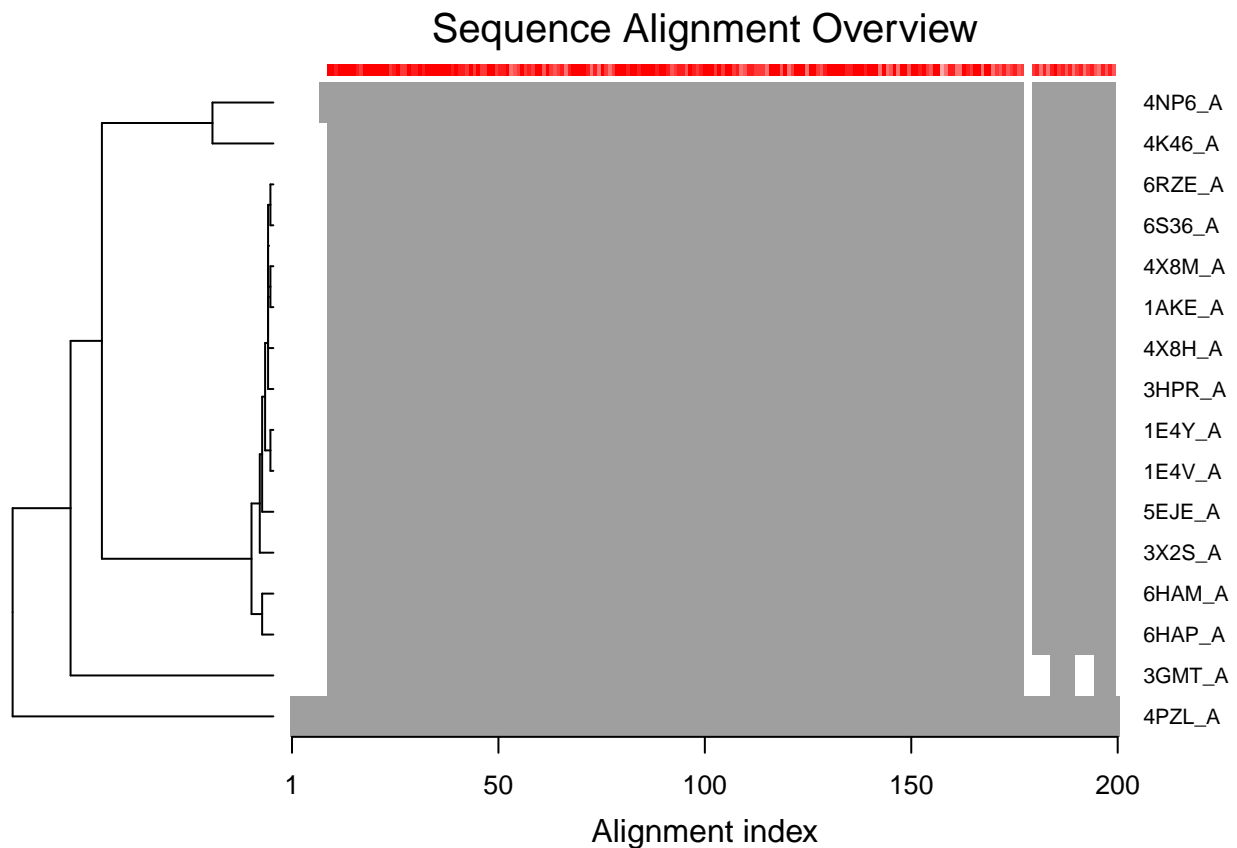

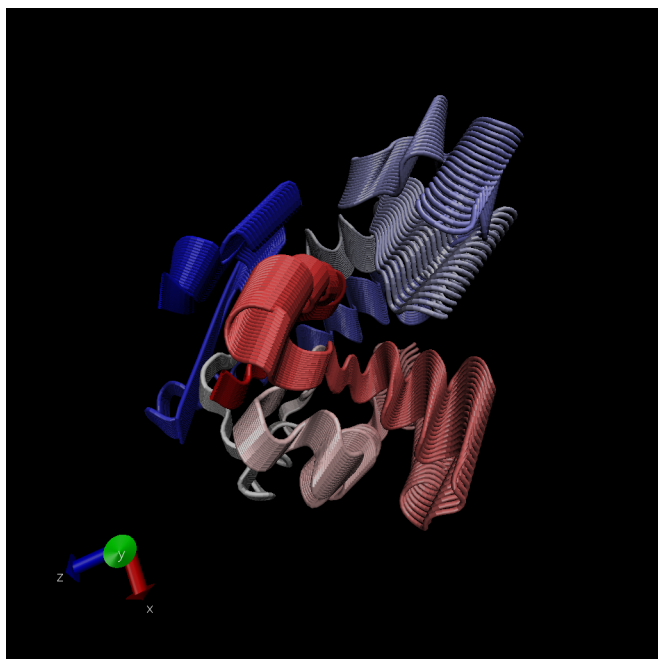we will use the bio3d pca() function which is designed for protein structure data.

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

make a trajectory vsiualization of the motion captured by the first Principal component

Visualize first principal component

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

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

MSA

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

devtools::install_bitbucket("Grantlab/bio3d-view")

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

TRUE

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

214