

Class12: Structural Bioinformatics II

Barry (PID: 911)

11/4/2021

Comparative analysis of protein structures

Using the bio3d package.

```
library(bio3d)

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

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

```
pdb
```

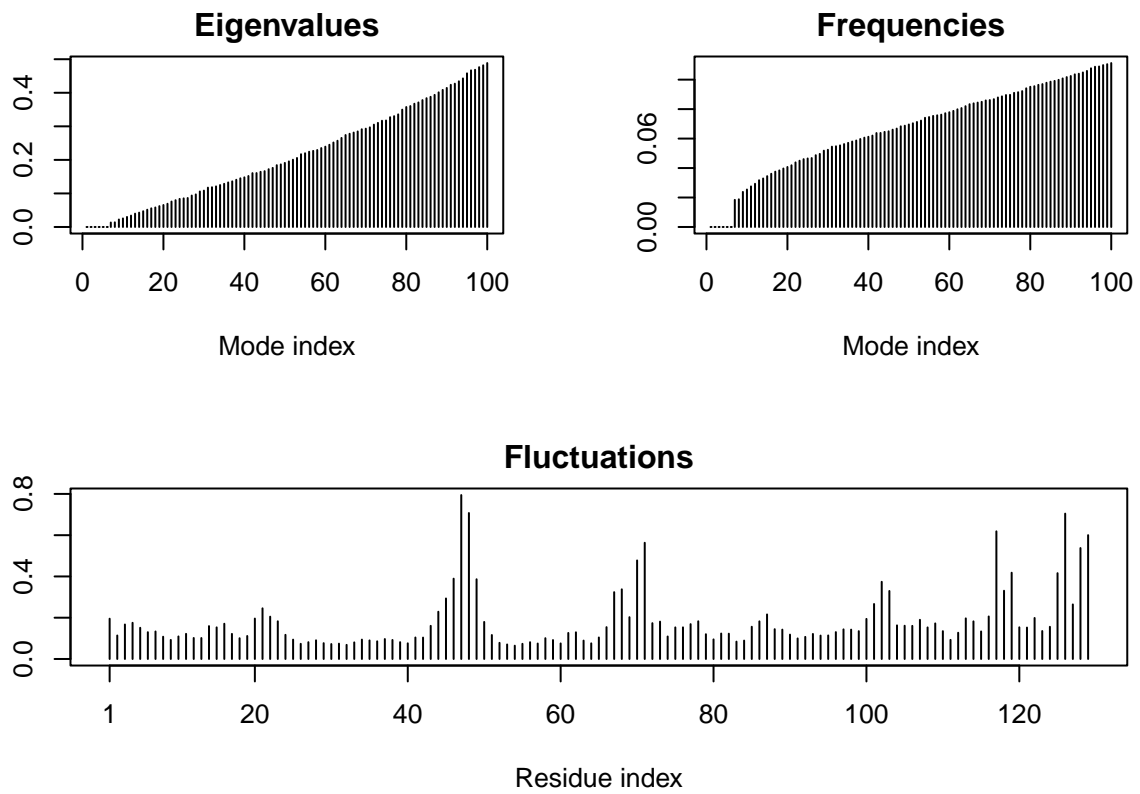
```
##
## Call: read.pdb(file = "1hel")
##
## Total Models#: 1
## Total Atoms#: 1186, XYZs#: 3558 Chains#: 1 (values: A)
##
## Protein Atoms#: 1001 (residues/Calpha atoms#: 129)
## Nucleic acid Atoms#: 0 (residues/phosphate atoms#: 0)
##
## Non-protein/nucleic Atoms#: 185 (residues: 185)
## Non-protein/nucleic resid values: [ HOH (185) ]
##
## Protein sequence:
## KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINS
## RWWCNDGRTPGSRNLCNIPCSALLSSDITASVNC AKKIVSDGNGMNAWVAWRNRCKGTDV
## QAWIRGCRL
##
## + attr: atom, xyz, seqres, helix, sheet,
## calpha, remark, call
```

Let's use a bioinformatics method called NMA (Normal Mode Analysis) to predict the dynamics (flexibility) of this enzyme.

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

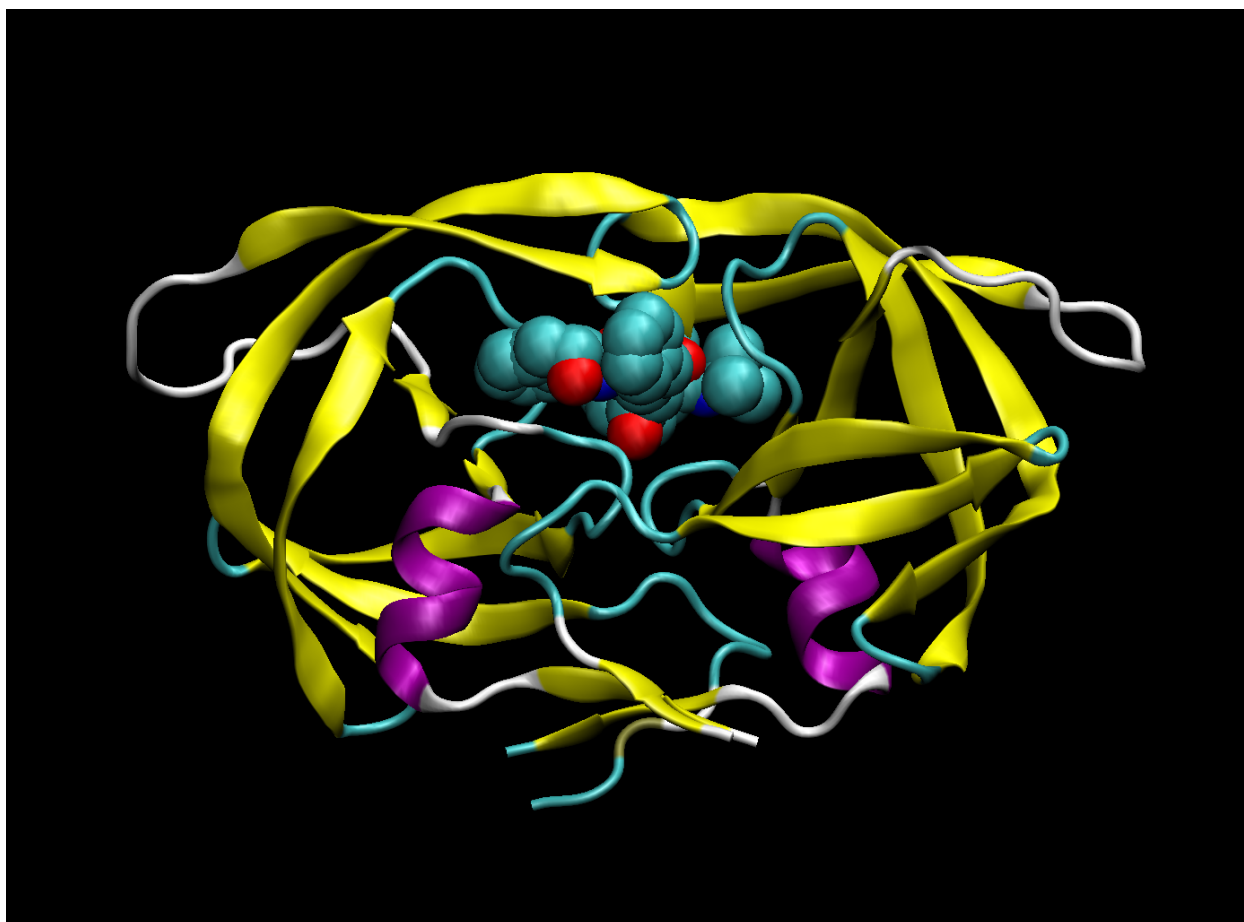
```
## Building Hessian... Done in 0.007 seconds.
## Diagonalizing Hessian... Done in 0.063 seconds.
```

```
plot(modes)
```



Make a “move” of its predicted motion. We often call this a “trajectory”.

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



Analysis of ADK

```
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  MRIILLGAPGACKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLV
##          1          .          .          .          .          60
##
##          61          .          .          .          .          120
## pdb|1AKE|A  DELVIALVKERIAQEDCRNGFLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVDRI
##          61          .          .          .          .          120
##
##          121         .          .          .          .          180
## pdb|1AKE|A  VGRRVHAPSGRVYHVKNPPKVEGKDDVTGEELTRKDDQEETVRKRLVEYHQMTAPLIG
```

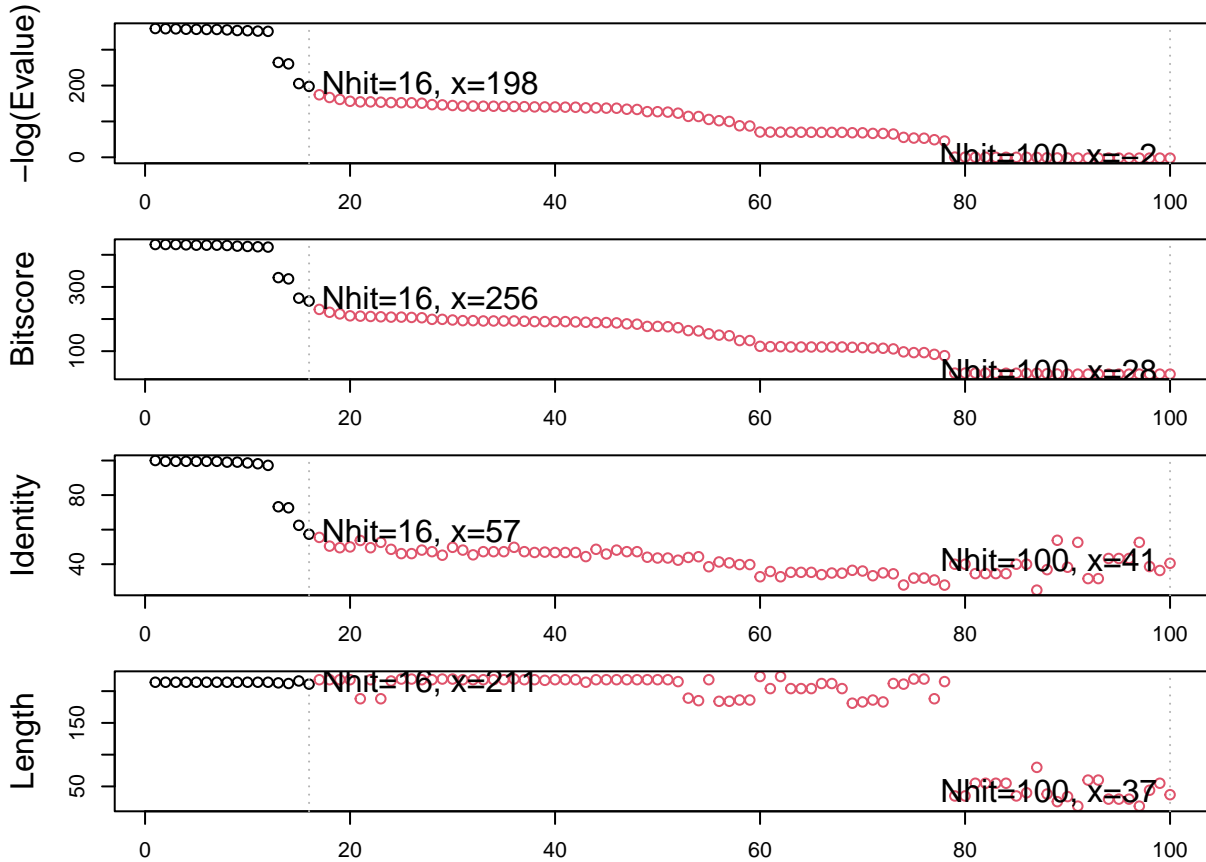
```
##          121          .          .          .          .          .          180
##
##          181          .          .          .          .          .          214
## pdb|1AKE|A  YYSKEAEAGNTKYAKVDGTPVAEVRADLEKILG
##          181          .          .          .          .          .          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
```

```
# Run BLAST from BLAST
#blast <- blast.pdb(aa)
```

```
load("mydata.RData")
```

```
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"
```

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

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
```

```

## pdb/split_chain/6HAM_A.pdb
## pdb/split_chain/4K46_A.pdb
## pdb/split_chain/4NP6_A.pdb
## pdb/split_chain/3GMT_A.pdb
## pdb/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: pdb/split_chain/1AKE_A.pdb
##   PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 2   name: pdb/split_chain/4X8M_A.pdb
## pdb/seq: 3   name: pdb/split_chain/6S36_A.pdb
##   PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 4   name: pdb/split_chain/6RZE_A.pdb
##   PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 5   name: pdb/split_chain/4X8H_A.pdb
## pdb/seq: 6   name: pdb/split_chain/3HPR_A.pdb
##   PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 7   name: pdb/split_chain/1E4V_A.pdb
## pdb/seq: 8   name: pdb/split_chain/5EJE_A.pdb
##   PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 9   name: pdb/split_chain/1E4Y_A.pdb
## pdb/seq: 10  name: pdb/split_chain/3X2S_A.pdb
## pdb/seq: 11  name: pdb/split_chain/6HAP_A.pdb
## pdb/seq: 12  name: pdb/split_chain/6HAM_A.pdb
##   PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 13  name: pdb/split_chain/4K46_A.pdb
##   PDB has ALT records, taking A only, rm.alt=TRUE
## pdb/seq: 14  name: pdb/split_chain/4NP6_A.pdb
## pdb/seq: 15  name: pdb/split_chain/3GMT_A.pdb
## pdb/seq: 16  name: pdb/split_chain/4PZL_A.pdb

```

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 TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
## [Truncated_Name:2] 4X8M_A.pdb TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
## [Truncated_Name:3] 6S36_A.pdb TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
## [Truncated_Name:4] 6RZE_A.pdb TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
## [Truncated_Name:5] 4X8H_A.pdb TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
## [Truncated_Name:6] 3HPR_A.pdb TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
## [Truncated_Name:7] 1E4V_A.pdb TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
## [Truncated_Name:8] 5EJE_A.pdb TGDMLRAAVKSGSELGKQAKDIMDACKLVDELVIALVKE
## [Truncated_Name:9] 1E4Y_A.pdb TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
## [Truncated_Name:10] 3X2S_A.pdb TGDMLRAAVKSGSELGKQAKDIMDCGKLVDELVIALVKE
## [Truncated_Name:11] 6HAP_A.pdb TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVRE
## [Truncated_Name:12] 6HAM_A.pdb TGDMLRAAIKSGSELGKQAKDIMDAGKLVDEIIIALVKE
## [Truncated_Name:13] 4K46_A.pdb TGDMLRAAIKAGTELGKQAKSVIDAGQLVSDDIILGLVKE
## [Truncated_Name:14] 4NP6_A.pdb TGDMLRAAIKAGTELGKQAKAVIDAGQLVSDDIILGLIKE
## [Truncated_Name:15] 3GMT_A.pdb TGDMLRAAVKAGTPLGVEAKTYMDEGKLVPSDLIIGLVKE
## [Truncated_Name:16] 4PZL_A.pdb TGDMIRETIKSGSALGQELKKVLDAGELVSDEFI IKIVKD
##
##          ****~*  ^*  ^*  **  *  ^*  **  *  ^^  ~~~~
##          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 RLKEADCANGYLFDFGFPRTIAQADAMKEAGVAIDYVLEID
## [Truncated_Name:16] 4PZL_A.pdb RISKNCNNGFLLDGVPRTIPQAQELDKLGVNIDYIVEVD
##
##          *~  *  *~*  **  *****  **  ^  *~  ^~~~*  *
##          81          .          .          .          120
##
##          121          .          .          .          160
## [Truncated_Name:1] 1AKE_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPVKVEGKDDVTG
## [Truncated_Name:2] 4X8M_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPVKVEGKDDVTG
## [Truncated_Name:3] 6S36_A.pdb VPDELIVDKIVGRRVHAPSGRVYHVKFNPVKVEGKDDVTG
## [Truncated_Name:4] 6RZE_A.pdb VPDELIVDAIVGRRVHAPSGRVYHVKFNPVKVEGKDDVTG
## [Truncated_Name:5] 4X8H_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPVKVEGKDDVTG

```

```

## [Truncated_Name:6] 3HPR_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDGTG
## [Truncated_Name:7] 1E4V_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
## [Truncated_Name:8] 5EJE_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
## [Truncated_Name:9] 1E4Y_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
## [Truncated_Name:10] 3X2S_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
## [Truncated_Name:11] 6HAP_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
## [Truncated_Name:12] 6HAM_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
## [Truncated_Name:13] 4K46_A.pdb VADSVIVERMAGRAHLASGRTYHNVYNPPKVEGKDDVTG
## [Truncated_Name:14] 4NP6_A.pdb VADDVIVERMAGRAHLPSGRTYHVYVYNPPKVEGKDDVTG
## [Truncated_Name:15] 3GMT_A.pdb VPFSEIIERMSGRRTHPASGRTYHVKNPPKVEGKDDVTG
## [Truncated_Name:16] 4PZL_A.pdb VADNLLIERITGRIHPASGRTYHTKFNPPKVADKDDVTG
##
## * ^^^ ^ *** * *** * ^***** ** **
## 121 . . . 160
##
## 161 . . 200
## [Truncated_Name:1] 1AKE_A.pdb EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
## [Truncated_Name:2] 4X8M_A.pdb EELTTRKDDQEETVRKRLVEWHQM TAPLIGYYSKEAEAGN
## [Truncated_Name:3] 6S36_A.pdb EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
## [Truncated_Name:4] 6RZE_A.pdb EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
## [Truncated_Name:5] 4X8H_A.pdb EELTTRKDDQEETVRKRLVEYHQM TAA LIGYYSKEAEAGN
## [Truncated_Name:6] 3HPR_A.pdb EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
## [Truncated_Name:7] 1E4V_A.pdb EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
## [Truncated_Name:8] 5EJE_A.pdb EELTTRKDDQEECVRKRLVEYHQM TAPLIGYYSKEAEAGN
## [Truncated_Name:9] 1E4Y_A.pdb EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
## [Truncated_Name:10] 3X2S_A.pdb EELTTRKDDQEETVRKRLCEYHQM TAPLIGYYSKEAEAGN
## [Truncated_Name:11] 6HAP_A.pdb EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
## [Truncated_Name:12] 6HAM_A.pdb EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
## [Truncated_Name:13] 4K46_A.pdb EDLVIREDDKEETVLARLGVYHNQTAPLIA YYGKEAEAGN
## [Truncated_Name:14] 4NP6_A.pdb EDLVIREDDKEETVRARLNVYHTQTAPLIE YYGKEAAAGK
## [Truncated_Name:15] 3GMT_A.pdb EPLVQRDDDKKEETVKKRLDVYEAQTKPLITY YGDWARRGA
## [Truncated_Name:16] 4PZL_A.pdb EPLITRTDDNEDTVKQRLSVYHAQTAKLIDFYR NFSSTNT
##
## * * * ** * ^ * ** ^ * ** ^*
## 161 . . . 200
##
## 201 . . 227
## [Truncated_Name:1] 1AKE_A.pdb T--KYAKVDG TKPVAEVRADLEKILG-
## [Truncated_Name:2] 4X8M_A.pdb T--KYAKVDG TKPVAEVRADLEKILG-
## [Truncated_Name:3] 6S36_A.pdb T--KYAKVDG TKPVAEVRADLEKILG-
## [Truncated_Name:4] 6RZE_A.pdb T--KYAKVDG TKPVAEVRADLEKILG-
## [Truncated_Name:5] 4X8H_A.pdb T--KYAKVDG TKPVAEVRADLEKILG-
## [Truncated_Name:6] 3HPR_A.pdb T--KYAKVDG TKPVAEVRADLEKILG-
## [Truncated_Name:7] 1E4V_A.pdb T--KYAKVDG TKPVAEVRADLEKILG-
## [Truncated_Name:8] 5EJE_A.pdb T--KYAKVDG TKPVAEVRADLEKILG-
## [Truncated_Name:9] 1E4Y_A.pdb T--KYAKVDG TKPVAEVRADLEKILG-
## [Truncated_Name:10] 3X2S_A.pdb T--KYAKVDG TKPVAEVRADLEKILG-
## [Truncated_Name:11] 6HAP_A.pdb T--KYAKVDG TKPVCEVRADLEKILG-
## [Truncated_Name:12] 6HAM_A.pdb T--KYAKVDG TKPVCEVRADLEKILG-
## [Truncated_Name:13] 4K46_A.pdb T--QYLKFDG TKAVAEVSAELEKALA-
## [Truncated_Name:14] 4NP6_A.pdb T--QYLKFDG TKQVSEVSADIAKALA-
## [Truncated_Name:15] 3GMT_A.pdb E-----NGLKAPA-----YRKISG-
## [Truncated_Name:16] 4PZL_A.pdb KIPKYIKINGDQAVEKVSQDIFDQLNK
##
## *
## 201 . . 227
##

```

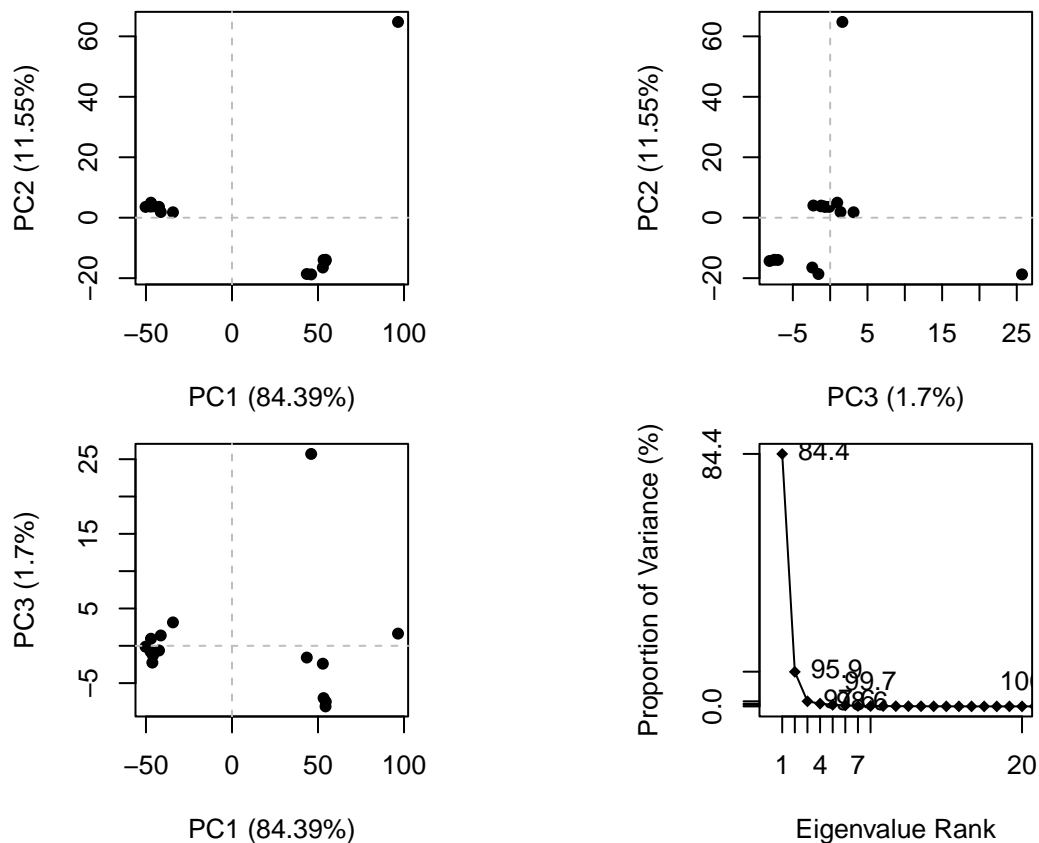


```
## Call:
##   pdbaln(files = files, fit = TRUE)
##
## Class:
##   pdba, fasta
##
## Alignment dimensions:
##   16 sequence rows; 227 position columns (204 non-gap, 23 gap)
##
## + attr: xyz, resno, b, chain, id, ali, resid, sse, call
```

PCA

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

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



Make a trajectory visualization of the the motion captured by the first Principal Component

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

