

Class11: Structural Bioinformatics Part 2

Camryn McCann (PID: A15437387)

11/4/2021

#Comparative Structure Analysis of Adenylate Kinase

Using 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
```

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

There are 129 amino acid residues in the pdb object 1hel.

Q8: Name one of the non-protein residues?

HOH

Q9: How many protein chains are in this structure?

1 protein chain in this structure

#Comparative Structure Analysis of Adenylate Kinase

First we must install the following packages in the console if we do not already have them: bio3d, ggplot2, ggrepel, devtools, BiocManager, msa, and bitbucket “Grantlab/bio3d-view”

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

msa was only found on BioConductor

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

The package “Grantlab/bio3d-view” is not found on BioConductor or CRAN, it could only be downloaded from BitBucket, using the devtools function of 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?

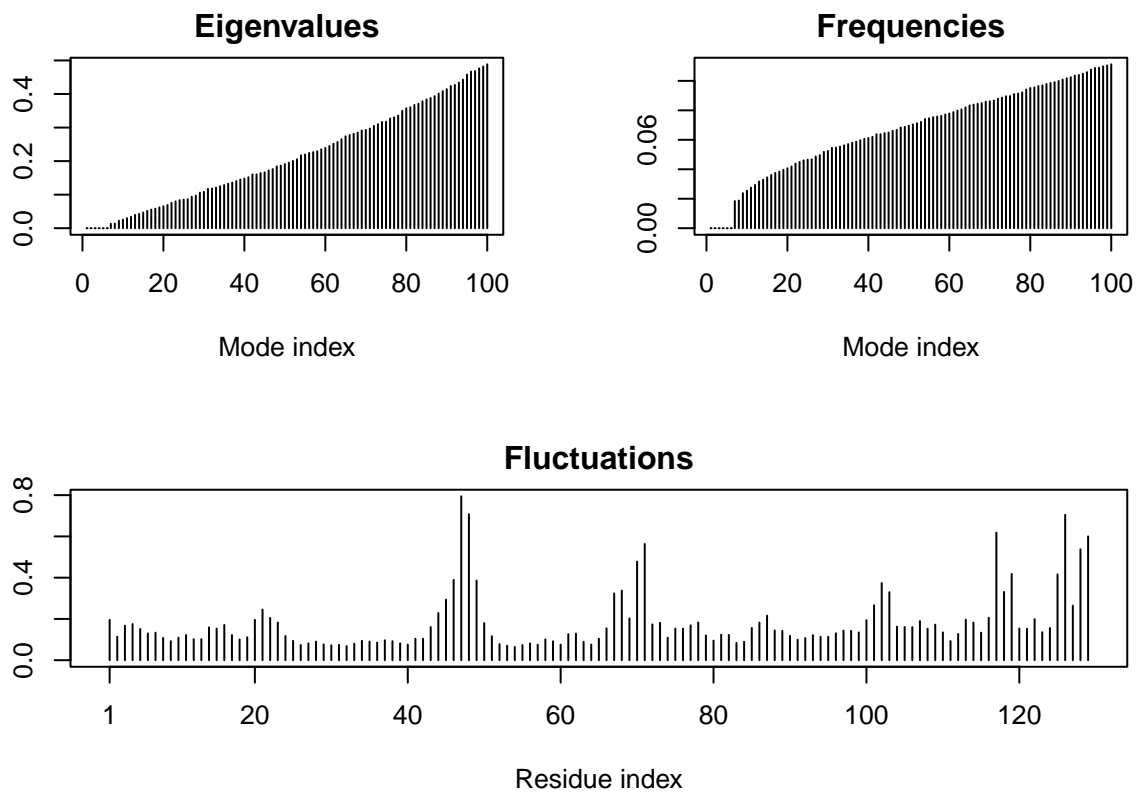
TRUE

Let's use a bioinformatics method called NMA(Normal Mode Analysis) to predict the dynamics (flexibility) of this enzyme (*from here down was all done in class on Thursday 11/4 with professor*)

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

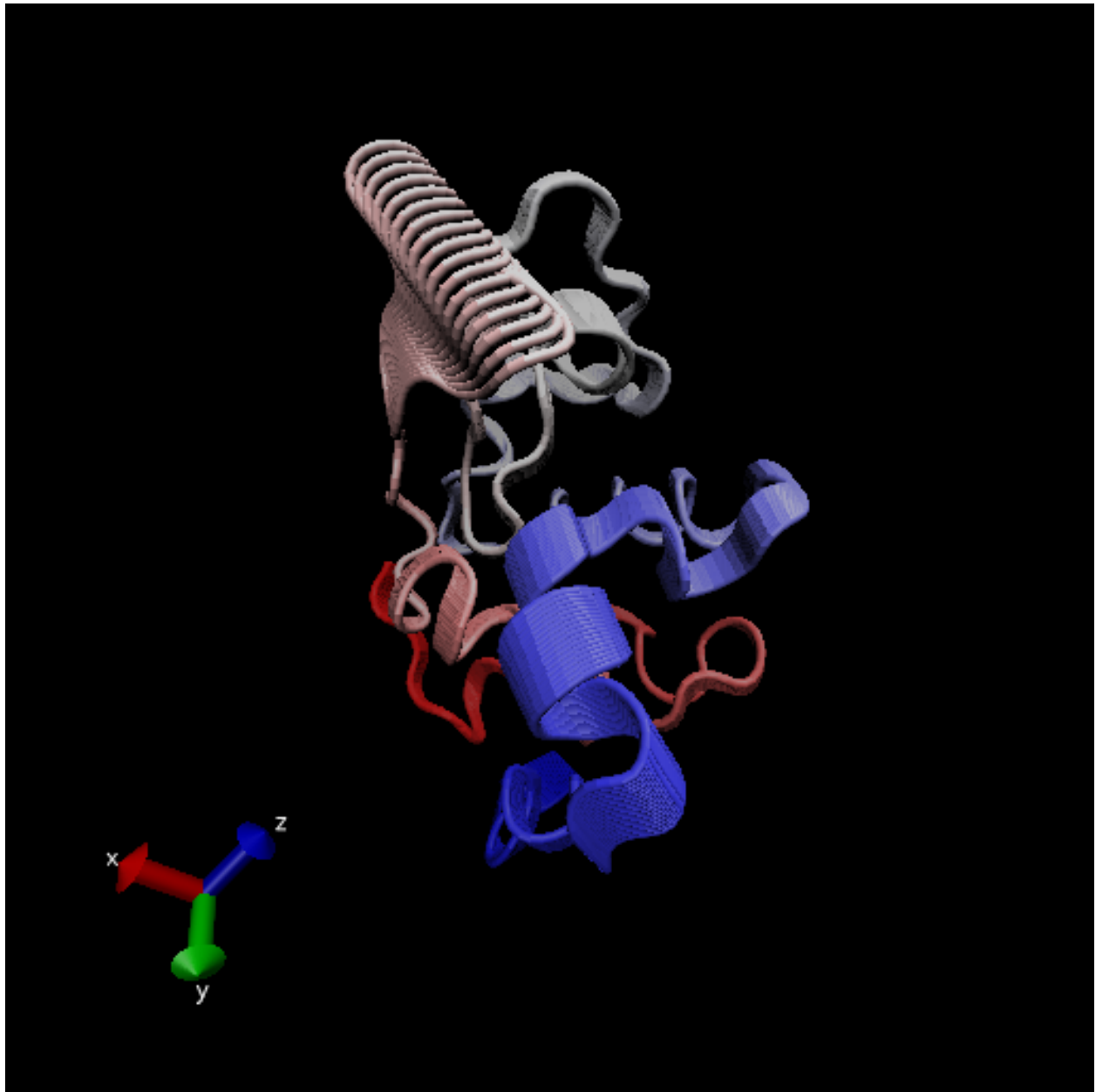
```
## Building Hessian... Done in 0.058 seconds.  
## Diagonalizing Hessian... Done in 0.176 seconds.
```

```
plot(modes)
```



Make a trajectory! (movie of its predicted motion).

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

214

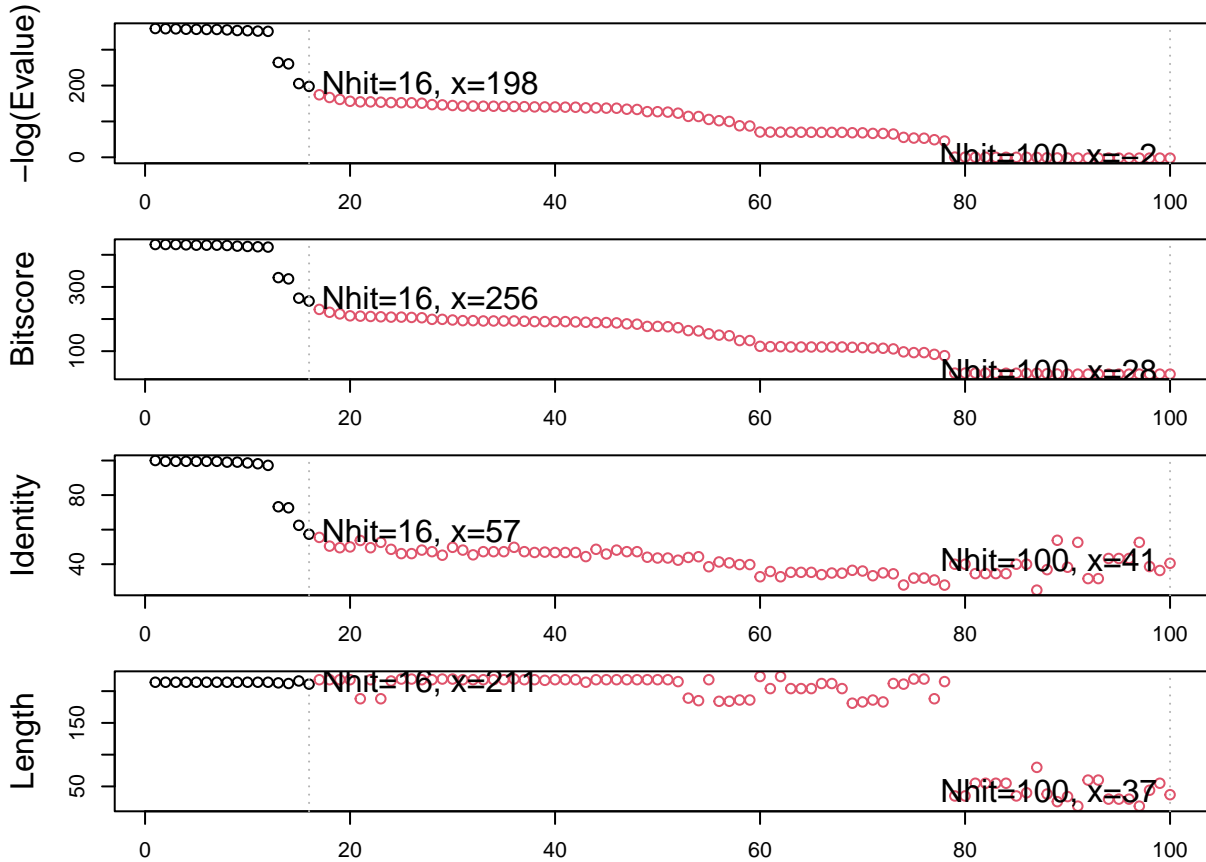
```
#Run BLAST
```

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

```
## Searching ... please wait (updates every 5 seconds) RID = SHADG6ZZ013
## .
## Reporting 100 hits
```

```
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 relate PDB files
```

```
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

##      |
```

Now, let's do a multiple structure alignment

```
pdbs <- 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
```

```

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

```

pdbc

```

##                                     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 TGDMLRAAIKAGTELKGQAKSVIDAGQLVSDDIILGLVKE
## [Truncated_Name:14] 4NP6_A.pdb TGDMLRAAIKAGTELKGQAKAVIDAGQLVSDDIILGLIKE
## [Truncated_Name:15] 3GMT_A.pdb TGDMLRAAVKAGTPLGVEAKTYMDEGKLPDSLIIIGLVKE
## [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

```

```

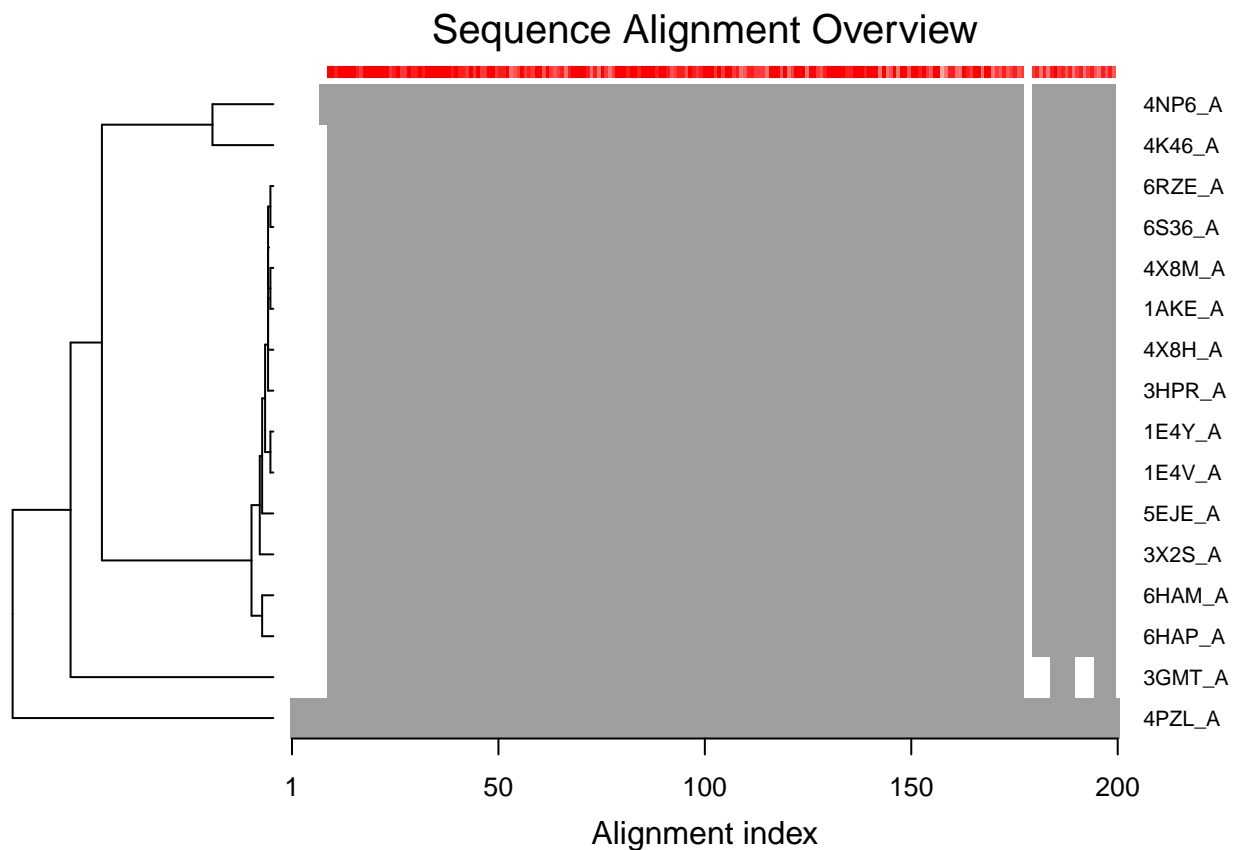
## [Truncated_Name:5] 4X8H_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
## [Truncated_Name:6] 3HPR_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDGTG
## [Truncated_Name:7] 1E4V_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
## [Truncated_Name:8] 5EJE_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
## [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      VADSVIVERMAGRRAHLASGRTYHNVYNPPKVEGKDDVTG
## [Truncated_Name:14] 4NP6_A.pdb      VADDVIVERMAGRRAHLPSGRTYHVYNPPKVEGKDDVTG
## [Truncated_Name:15] 3GMT_A.pdb      VPFSEIIERMSGRRTHPASGRTYHVKNPPKVEGKDDVTG
## [Truncated_Name:16] 4PZL_A.pdb      VADNLLIERITGRIHPASGRTYHTKFNPPKVADKDDVTG
##                                     *   ^^^ ^ *** * *** * ^***** ** **
##                                     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      EPLVQRDDDKKEETVKKRLDVYEAQTKPLITYYGDWARRGA
## [Truncated_Name:16] 4PZL_A.pdb      EPLITRTDDNEDTVKQRLSVYHAQTAKLIDFYRNFSSNT
##                                     * * * ** * ^ * ** ^ * ** ^*
##                                     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--QYLKFDGTKQVSEVSADIKALA-
## [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
```

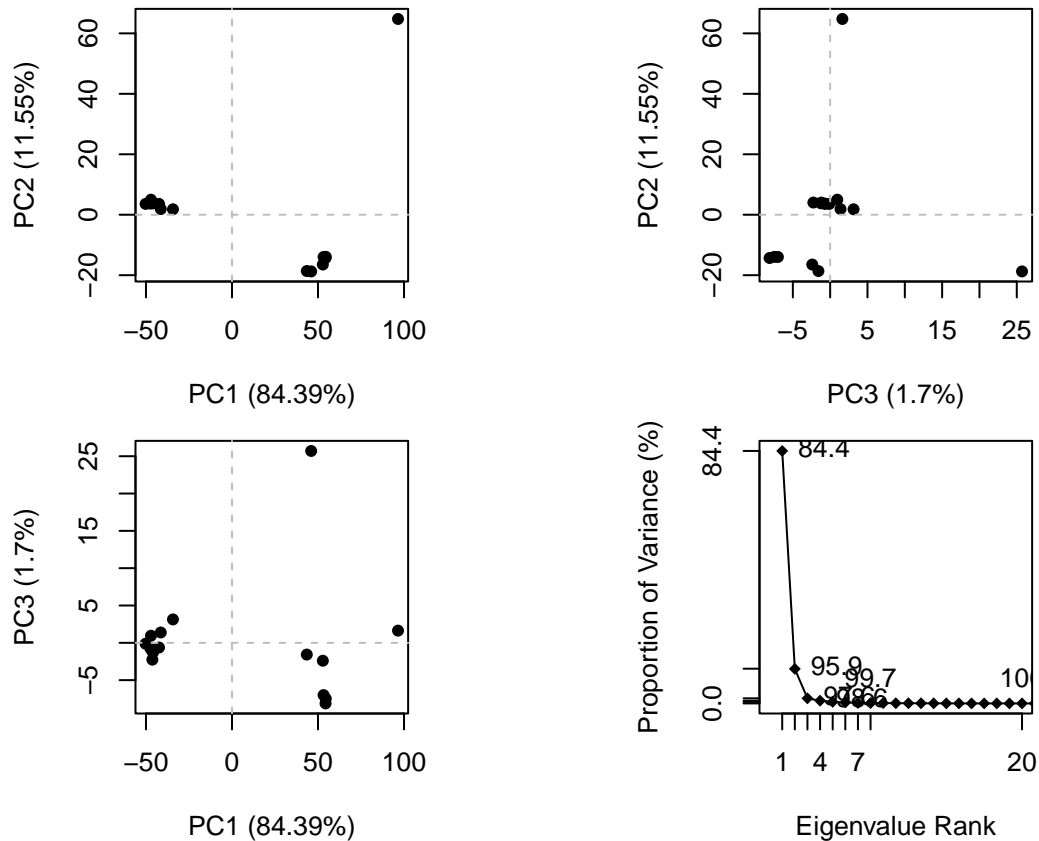
```
# Vector containing PDB codes for figure axis
ids <- basename.pdb(pdba$id)

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



Let's do PCA!

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



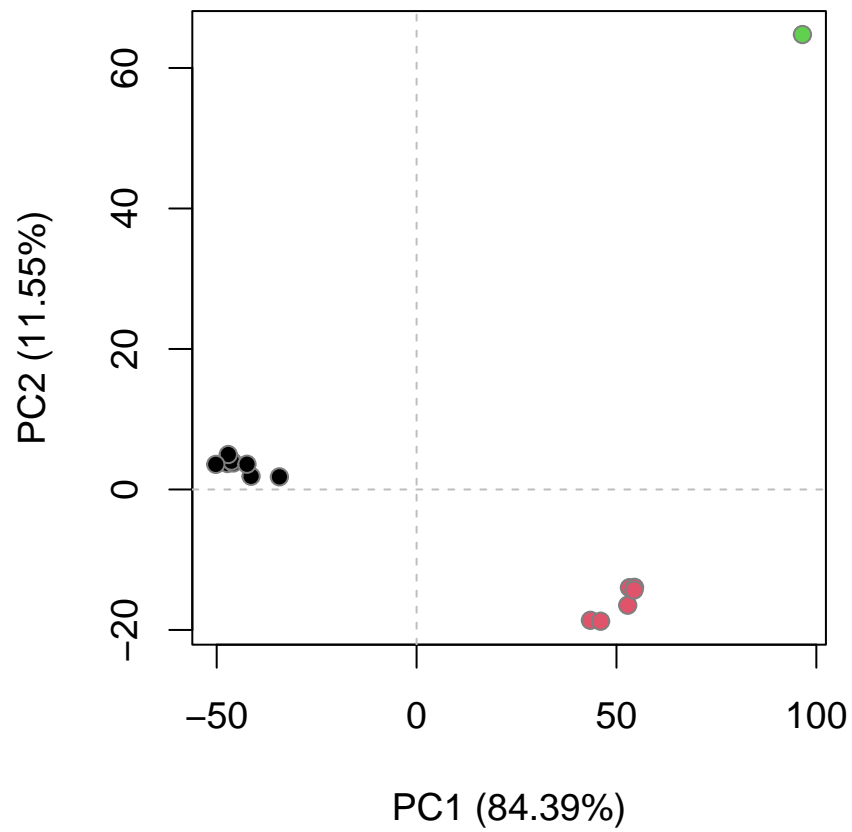
Function `rmsd()` will calculate all pairwise RMSD values of the structural ensemble. This facilitates clustering analysis based on the pairwise structural deviation.

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

```
## Warning in rmsd(pdb): 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(pc.xray, 1:2, col="grey50", bg=grps.rd, pch=21, cex=1)
```



Next, we can make a trajectory visualization of the motion captured by the first Principal Component.

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

We use the file on VMD and can visualize the movement.

