

# Class 10 : Comparative structure analysis of Adenylate Kinase

AUTHOR

Loreen A17059289

## Comparative Analysis of ADK

We will search the entire PDB for related structures using BLAST, fetch, align and superpose the identified structures, perform PCA and finally calculate the normal modes of each individual structure in order to probe for potential differences in structural flexibility.

Questions 10-12:

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?:

bio3d-view

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

TRUE

```
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
MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDAGKLV

```

```

      1      .      .      .      .
.      60

```

```

      61      .      .      .      .
.      120
pdb|1AKE|A
DELVIALVKERIAQEDCRNGFLLDGFRTIPQADAMKEAGINVDYVLEFDVPDELIVDRI

```

```

      61      .      .      .      .
.      120

```

```

      121      .      .      .      .
.      180
pdb|1AKE|A
VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG

```

```

      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

We can now run BLAST with this sequence:

```
# Blast or hmmer search
# b <- blast.pdb(aa)
```

```
# hits <- plot(b)
```

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

214

Let's see what is in our `hits` object. Use `pdb.id` to access them.

```
# hits$ pdb.id

hits <- NULL
hits$ pdb.id <- c('1AKE_A', '6S36_A', '6RZE_A', '3HPR_A', '1E4V_A', '1E4V_A')
```

Now we can download all of these PDB structures:

```
# Download related 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/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/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):
```

pdb/5EJE.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbc", split = TRUE,  
gzip = TRUE):

pdb/1E4Y.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbc", split = TRUE,  
gzip = TRUE):

pdb/3X2S.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbc", split = TRUE,  
gzip = TRUE):

pdb/6HAP.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbc", split = TRUE,  
gzip = TRUE):

pdb/6HAM.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbc", split = TRUE,  
gzip = TRUE):

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

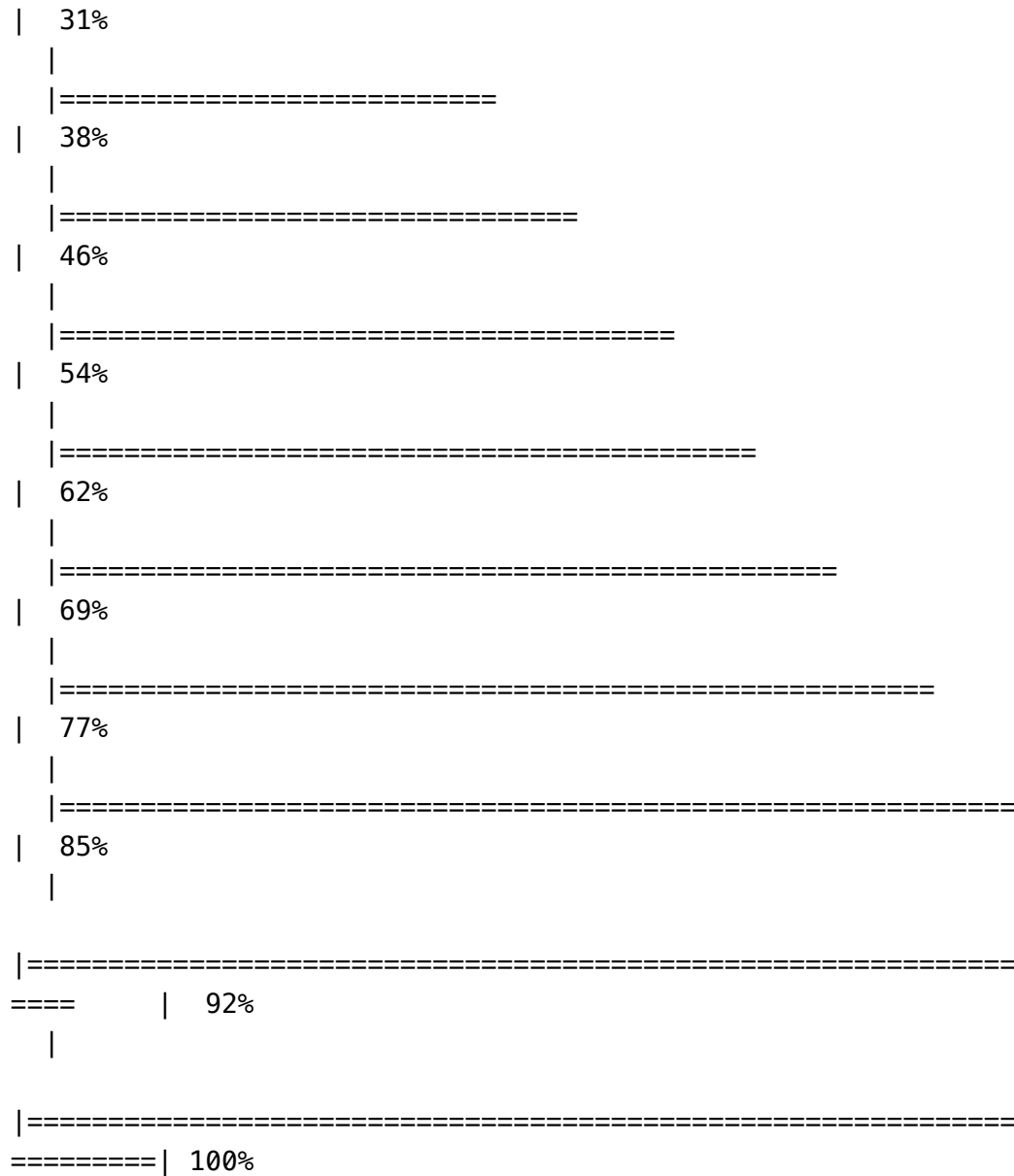
Warning in get.pdb(hits\$pdb.id, path = "pdbc", split = TRUE,  
gzip = TRUE):

pdb/3GMT.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbc", split = TRUE,  
gzip = TRUE):

pdb/4PZL.pdb.gz exists. Skipping download

```
|
|
| 0%
|
|=====
| 8%
|
|=====
| 15%
|
|=====
| 23%
|
|=====
```



Now I want to align and superpose these structures which are all over the place.

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

Reading PDB files:

```
pdbs/split_chain/1AKE_A.pdb
pdbs/split_chain/6S36_A.pdb
pdbs/split_chain/6RZE_A.pdb
pdbs/split_chain/3HPR_A.pdb
pdbs/split_chain/1E4V_A.pdb
pdbs/split_chain/5EJE_A.pdb
```

```

pdb/split_chain/1E4Y_A.pdb
pdb/split_chain/3X2S_A.pdb
pdb/split_chain/6HAP_A.pdb
pdb/split_chain/6HAM_A.pdb
pdb/split_chain/4K46_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/6S36_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 3    name: pdb/split_chain/6RZE_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 4    name: pdb/split_chain/3HPR_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 5    name: pdb/split_chain/1E4V_A.pdb
pdb/seq: 6    name: pdb/split_chain/5EJE_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 7    name: pdb/split_chain/1E4Y_A.pdb
pdb/seq: 8    name: pdb/split_chain/3X2S_A.pdb
pdb/seq: 9    name: pdb/split_chain/6HAP_A.pdb
pdb/seq: 10   name: pdb/split_chain/6HAM_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 11   name: pdb/split_chain/4K46_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 12   name: pdb/split_chain/3GMT_A.pdb
pdb/seq: 13   name: pdb/split_chain/4PZL_A.pdb

```

Let's look at what `pdb` looks like:

```

# Alignment of all structures
pdb

```

```

1      .      .      .
40
[Truncated_Name:1]1AKE_A.pdb -----
MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:2]6S36_A.pdb -----
MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:3]6RZE_A.pdb -----
MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:4]3HPR_A.pdb -----
MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:5]1E4V_A.pdb -----
MRIILLGAPVAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:6]5EJE_A.pdb -----
MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:7]1E4Y_A.pdb -----
MRIILLGALVAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:8]3X2S_A.pdb -----
MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:9]6HAP_A.pdb -----
MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:10]6HAM_A.pdb -----
MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:11]4K46_A.pdb -----
MRIILLGAPGAGKGTQAQFIMAKFGIPQIS
[Truncated_Name:12]3GMT_A.pdb -----
MRLILLGAPGAGKGTQANFIKEKFGIPQIS
[Truncated_Name:13]4PZL_A.pdb
TENLYFQSNAMRIILLGAPGAGKGTQAKIIEQKYNIAHIS
**^*****  *****  *
*^ *  **

```

```

1      .      .      .
40
41      .      .      .
80
[Truncated_Name:1]1AKE_A.pdb
TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE
[Truncated_Name:2]6S36_A.pdb
TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE
[Truncated_Name:3]6RZE_A.pdb
TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE
[Truncated_Name:4]3HPR_A.pdb
TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE
[Truncated_Name:5]1E4V_A.pdb
TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE

```

```

[Truncated_Name:6]5EJE_A.pdb
TGDMRLRAAVKSGSELGKQAKDIMDACKLVTDELVIALVKE
[Truncated_Name:7]1E4Y_A.pdb
TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE
[Truncated_Name:8]3X2S_A.pdb
TGDMRLRAAVKSGSELGKQAKDIMDCGKLVTDDELVIALVKE
[Truncated_Name:9]6HAP_A.pdb
TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVRE
[Truncated_Name:10]6HAM_A.pdb
TGDMRLRAAIKSGSELGKQAKDIMDAGKLVTDDEIIIALVKE
[Truncated_Name:11]4K46_A.pdb
TGDMRLRAAIKAGTELGKQAKSVIDAGQLVSDDIILGLVKE
[Truncated_Name:12]3GMT_A.pdb
TGDMRLRAAVKAGTPLGVEAKTYMDEGKLVDPDSLIIGLVKE
[Truncated_Name:13]4PZL_A.pdb
TGDMIRETIKSGSALGQELKKVLDAGELVSDEFIIKIVKD
                                     *****^*  ^* *^ **   *  ^*   **
*   ^^ ^*^^
                                     41           .           .           .
80
                                     81           .           .           .
120
[Truncated_Name:1]1AKE_A.pdb
RIAQEDCRNGFLLDGFRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:2]6S36_A.pdb
RIAQEDCRNGFLLDGFRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:3]6RZE_A.pdb
RIAQEDCRNGFLLDGFRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:4]3HPR_A.pdb
RIAQEDCRNGFLLDGFRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:5]1E4V_A.pdb
RIAQEDCRNGFLLDGFRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:6]5EJE_A.pdb
RIAQEDCRNGFLLDGFRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:7]1E4Y_A.pdb
RIAQEDCRNGFLLDGFRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:8]3X2S_A.pdb
RIAQEDSRNGFLLDGFRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:9]6HAP_A.pdb
RICQEDSRNGFLLDGFRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:10]6HAM_A.pdb
RICQEDSRNGFLLDGFRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:11]4K46_A.pdb
RIAQDDCAKGFLDGFRTIPQADGLKEVGVVVDYVIEFD

```



```

[Truncated_Name:12]3GMT_A.pdb
RLKEADCANGYLFDFGFPRTIAQADAMKEAGVAIDYVLEID
[Truncated_Name:13]4PZL_A.pdb
RISKNDCCNNGFLLDGVPRTIPQAQELDKLGVNIDYIVEVD
                                         *^  *  *^* ** ***** **  ^
*^  ^**^^*  *
                                         81      .      .      .
120
                                         121      .      .      .
160
[Truncated_Name:1]1AKE_A.pdb
VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:2]6S36_A.pdb
VPDELIVDKIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:3]6RZE_A.pdb
VPDELIVDAIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:4]3HPR_A.pdb
VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDGTG
[Truncated_Name:5]1E4V_A.pdb
VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:6]5EJE_A.pdb
VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:7]1E4Y_A.pdb
VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:8]3X2S_A.pdb
VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:9]6HAP_A.pdb
VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:10]6HAM_A.pdb
VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:11]4K46_A.pdb
VADSVIVERMAGRRRAHLASGRTYHNVNPPKVEGKDDVTG
[Truncated_Name:12]3GMT_A.pdb
VPFSEIIERMSGRRTHPASGRTYHVKFNPPKVEGKDDVTG
[Truncated_Name:13]4PZL_A.pdb
VADNLLIERITGRRIH PASGRTYHTKFNPPKVADKDDVTG
                                         *    ^^^ ^  *** *  *** **
^*****  *** **
                                         121      .      .      .
160
                                         161      .      .      .
200
[Truncated_Name:1]1AKE_A.pdb

```

```

EELTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:2]6S36_A.pdb
EELTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:3]6RZE_A.pdb
EELTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:4]3HPR_A.pdb
EELTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:5]1E4V_A.pdb
EELTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:6]5EJE_A.pdb
EELTRKDDQEECVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:7]1E4Y_A.pdb
EELTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:8]3X2S_A.pdb
EELTRKDDQEETVRKRLCEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:9]6HAP_A.pdb
EELTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:10]6HAM_A.pdb
EELTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:11]4K46_A.pdb
EDLVIREDDKEETVLARLGVYHNQTAPLIAYYGKEAEAGN
[Truncated_Name:12]3GMT_A.pdb
EPLVQRDDDKEETVKKRLDVYEAQTKPLITYYGDWARRGA
[Truncated_Name:13]4PZL_A.pdb
EPLITRTDDNEDTVKQRLSVYHAQTAKLIDFYRNFSSNTNT
                                * * * ** *^ * ** * * **
^*
                                161      .      .      .
200
                                201      .      .      227
[Truncated_Name:1]1AKE_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:2]6S36_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:3]6RZE_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:4]3HPR_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:5]1E4V_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:6]5EJE_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:7]1E4Y_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:8]3X2S_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:9]6HAP_A.pdb    T--KYAKVDGTKPVCEVRADLEKILG-
[Truncated_Name:10]6HAM_A.pdb    T--KYAKVDGTKPVCEVRADLEKILG-
[Truncated_Name:11]4K46_A.pdb    T--QYLKFDGTKA VA EVSAELEKALA-
[Truncated_Name:12]3GMT_A.pdb    E-----NGLKAPA-----YRKISG-
[Truncated_Name:13]4PZL_A.pdb    KIPKYIKINGDQAVEKVSQDIFDQLNK
                                *

```

201

227

Call:

```
pdbaln(files = files, fit = TRUE, exefile = "msa")
```

Class:

```
pdb, fasta
```

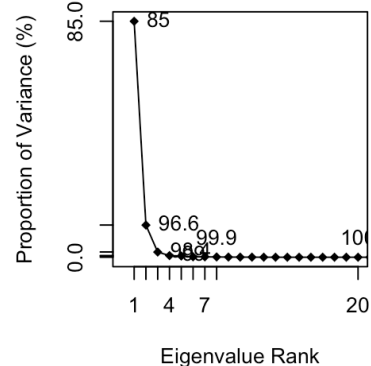
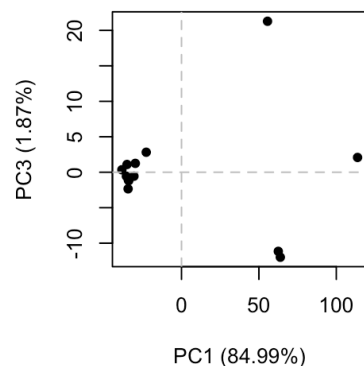
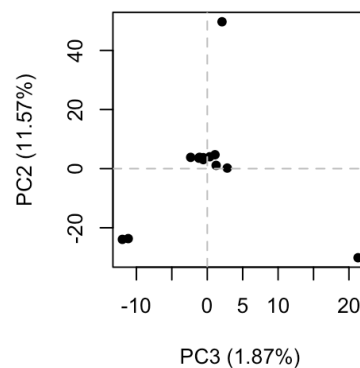
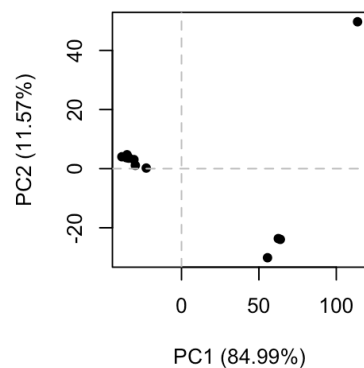
Alignment dimensions:

```
13 sequence rows; 227 position columns (204 non-gap, 23 gap)
```

```
+ attr: xyz, resno, b, chain, id, ali, resid, sse, call
```

Now we have our aligned and superposed structures which we can perform all sorts of analysis on. Let's do PCA:

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



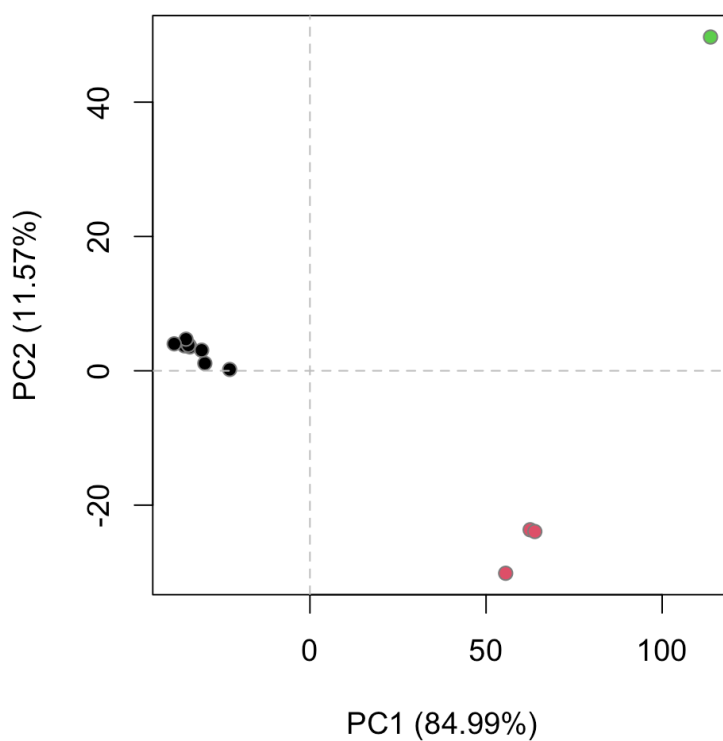
Results of PCA on Adenylate kinase X-ray structures. Each dot represents one PDB structure.

We can cluster the structures by RMSD(or any other method):

```
# 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)
```



OPTIONAL:

We can make a movie - also called a trajectory of the major differences(i.e. structural displacements) of ADK.

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

