

Class 10: Structural Bioinformatics (Part II)

Katherine Lim A15900881

Comparative analysis of ADK

Adenylate kinase (ADK) is an important drug target and we would love to know how it works (i.e. molecular mechanism).

There has been a lot of work done on this protein due to its importance including lots of crystal structures.

We will begin with getting an example ADK sequence from the database. We will then use this to find all ADK structures in the PDB.

```
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  MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLV
      1      .      .      .      .      .      60

      61      .      .      .      .      .      120
pdb|1AKE|A  DELVIALVKERIAQEDCRNGFLLDGFPR TIPQADAMKEAGINV DYVLEFDVPDELIVDRI
      61      .      .      .      .      .      120
```

```

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

```

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

# Plot a summary of search results
#hits <- plot(b)

hits <- NULL
hits$ pdb.id <- c('1AKE_A', '6S36_A', '6RZE_A', '3HPR_A', '1E4V_A', '5EJE_A', '1E4Y_A', '3X2S_A', '

```

Let's see what's in our hits object.

```
#hits$ pdb.id
```

Now we can download all these PDB structure files.

```

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

```

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

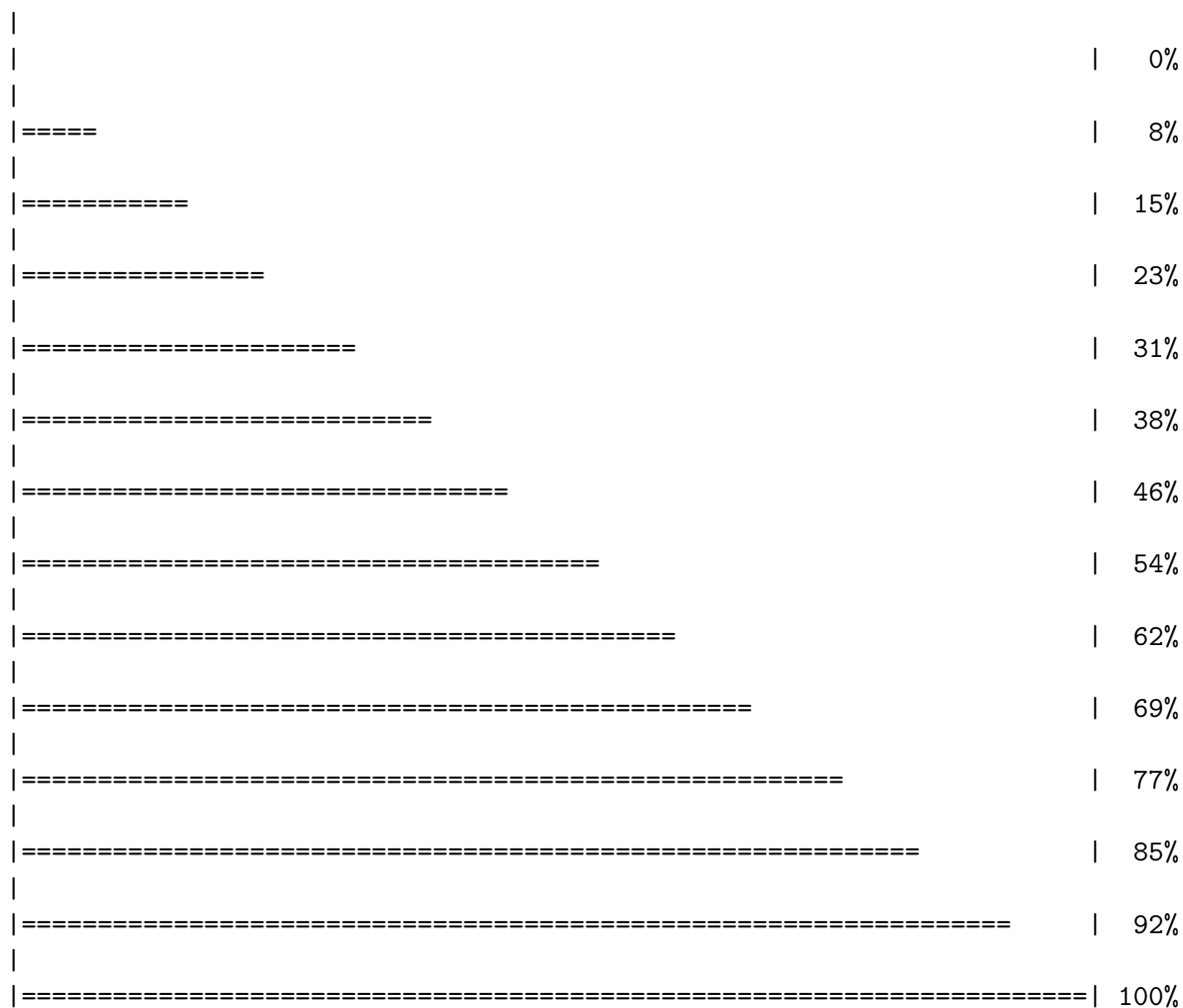
Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/6HAP.pdb.gz exists. Skipping download

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/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



files

```
[1] "pdbs/split_chain/1AKE_A.pdb" "pdbs/split_chain/6S36_A.pdb"
[3] "pdbs/split_chain/6RZE_A.pdb" "pdbs/split_chain/3HPR_A.pdb"
[5] "pdbs/split_chain/1E4V_A.pdb" "pdbs/split_chain/5EJE_A.pdb"
[7] "pdbs/split_chain/1E4Y_A.pdb" "pdbs/split_chain/3X2S_A.pdb"
[9] "pdbs/split_chain/6HAP_A.pdb" "pdbs/split_chain/6HAM_A.pdb"
[11] "pdbs/split_chain/4K46_A.pdb" "pdbs/split_chain/3GMT_A.pdb"
[13] "pdbs/split_chain/4PZL_A.pdb"
```

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
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/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/6S36_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 3  name: pdbs/split_chain/6RZE_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 4  name: pdbs/split_chain/3HPR_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 5  name: pdbs/split_chain/1E4V_A.pdb
pdb/seq: 6  name: pdbs/split_chain/5EJE_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 7  name: pdbs/split_chain/1E4Y_A.pdb
pdb/seq: 8  name: pdbs/split_chain/3X2S_A.pdb
pdb/seq: 9  name: pdbs/split_chain/6HAP_A.pdb
```

```

pdb/seq: 10  name: pdbs/split_chain/6HAM_A.pdb
PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 11  name: pdbs/split_chain/4K46_A.pdb
PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 12  name: pdbs/split_chain/3GMT_A.pdb
pdb/seq: 13  name: pdbs/split_chain/4PZL_A.pdb

```

Let's have a look at our pdbs object.

```
pdbs
```

```

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 TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
[Truncated_Name:2] 6S36_A.pdb TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
[Truncated_Name:3] 6RZE_A.pdb TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
[Truncated_Name:4] 3HPR_A.pdb TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
[Truncated_Name:5] 1E4V_A.pdb TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
[Truncated_Name:6] 5EJE_A.pdb TGDMLRAAVKSGSELGKQAKDIMDACKLVDELVIALVKE
[Truncated_Name:7] 1E4Y_A.pdb TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
[Truncated_Name:8] 3X2S_A.pdb TGDMLRAAVKSGSELGKQAKDIMDCGKLVDELVIALVKE
[Truncated_Name:9] 6HAP_A.pdb TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVRE
[Truncated_Name:10] 6HAM_A.pdb TGDMLRAAIIKSGSELGKQAKDIMDAGKLVDEIIIIALVKE
[Truncated_Name:11] 4K46_A.pdb TGDMLRAAIIKAGTELGKQAKSVIDAGQLVSDDIILGLVKE
[Truncated_Name:12] 3GMT_A.pdb TGDMLRAAVKAGTPLGVEAKTYMDEGKLVPSLIIGLVKE
[Truncated_Name:13] 4PZL_A.pdb TGDMIRETIKSGSALGQELKKVLDAGELVSDEFIIKIVKD

```

```

****~*  ~* *~ **  *  ~*  ** *  ^^ ~*~^
41      .      .      .      80

81      .      .      .      120
[Truncated_Name:1] 1AKE_A.pdb  RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:2] 6S36_A.pdb  RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:3] 6RZE_A.pdb  RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:4] 3HPR_A.pdb  RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:5] 1E4V_A.pdb  RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:6] 5EJE_A.pdb  RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:7] 1E4Y_A.pdb  RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:8] 3X2S_A.pdb  RIAQEDSRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:9] 6HAP_A.pdb  RICQEDSRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:10] 6HAM_A.pdb  RICQEDSRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:11] 4K46_A.pdb  RIAQDDCAKGFLLDGFPR TIPQADGLKEVG VVVVDYVIEFD
[Truncated_Name:12] 3GMT_A.pdb  RLKEADCANGYLF DGFPR TIAQADAMKEAGVAIDYVLEID
[Truncated_Name:13] 4PZL_A.pdb  RISKNDCNNGFLLDGVPR TIPQAQELDKLG VNI DYIVEVD
*~  *  *~* ** ***** **  ^  *~ ~**~*~* *
81      .      .      .      120

121     .      .      .      160
[Truncated_Name:1] 1AKE_A.pdb  VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:2] 6S36_A.pdb  VPDELIVDKIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:3] 6RZE_A.pdb  VPDELIVDAIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:4] 3HPR_A.pdb  VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDGTG
[Truncated_Name:5] 1E4V_A.pdb  VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:6] 5EJE_A.pdb  VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:7] 1E4Y_A.pdb  VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:8] 3X2S_A.pdb  VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:9] 6HAP_A.pdb  VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:10] 6HAM_A.pdb  VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:11] 4K46_A.pdb  VADSVIVERMAGRR AHLASGR TYHNVYNPPKVEGKDDVTG
[Truncated_Name:12] 3GMT_A.pdb  VPFSEIIERMSGRR THPASGR TYHV KFNPPKVEGKDDVTG
[Truncated_Name:13] 4PZL_A.pdb  VADNLLIERITGRRIHPASGR TYHTKFNPPKVADKDDVTG
*  ^^^ ^ *** *  *** **  ^***** *** **
121     .      .      .      160

161     .      .      .      200
[Truncated_Name:1] 1AKE_A.pdb  EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:2] 6S36_A.pdb  EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:3] 6RZE_A.pdb  EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:4] 3HPR_A.pdb  EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:5] 1E4V_A.pdb  EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN

```

```

[Truncated_Name:6] 5EJE_A.pdb      EELTTRKDDQEECVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:7] 1E4Y_A.pdb      EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:8] 3X2S_A.pdb      EELTTRKDDQEETVRKRLCEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:9] 6HAP_A.pdb      EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:10] 6HAM_A.pdb      EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:11] 4K46_A.pdb      EDLVIREDDKEETVLARLGVYHNQ TAPLIAYYGKEAEAGN
[Truncated_Name:12] 3GMT_A.pdb      EPLVQRDDDKKEETVKKRLDVYEA QTKPLITYYGDWARRGA
[Truncated_Name:13] 4PZL_A.pdb      EPLITRTDDNEDTVKQRLSVYHAQ TAKLIDFYRNFSSNTNT
                                     * * * * * ^ * * * * * ^ *
                                     161      .      .      .      200

                                     201      .      .      227
[Truncated_Name:1] 1AKE_A.pdb      T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:2] 6S36_A.pdb      T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:3] 6RZE_A.pdb      T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:4] 3HPR_A.pdb      T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:5] 1E4V_A.pdb      T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:6] 5EJE_A.pdb      T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:7] 1E4Y_A.pdb      T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:8] 3X2S_A.pdb      T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:9] 6HAP_A.pdb      T--KYAKVDGTPVCEVRADLEKILG-
[Truncated_Name:10] 6HAM_A.pdb      T--KYAKVDGTPVCEVRADLEKILG-
[Truncated_Name:11] 4K46_A.pdb      T--QYLKFDGTPKVAEVS AELEKALA-
[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:

```
pdbs, fasta
```

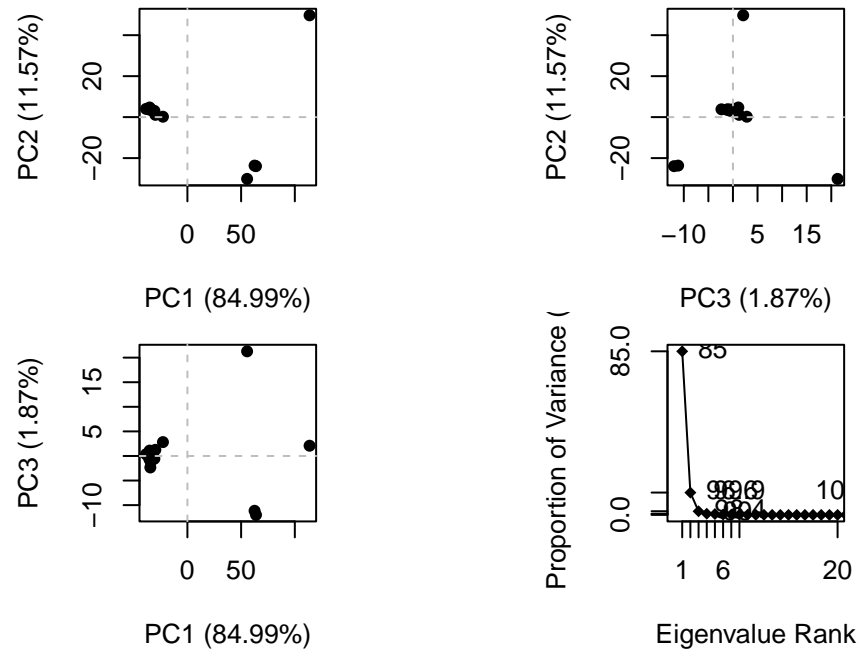
Alignment dimensions:

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

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

Now that we have our aligned and superposed structures we can perform all sorts of analyses on them. Let's first do PCA.


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



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

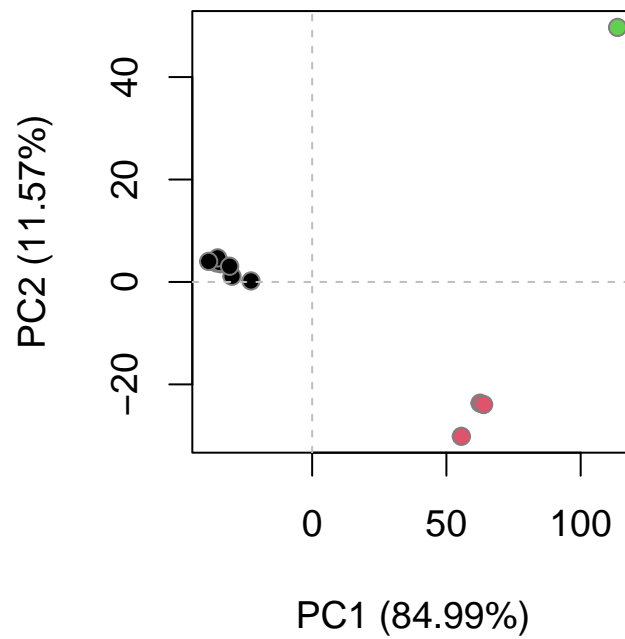
We can cluster the structures by RMSD (or any other method).

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

Warning in rmsd(pdbbs): 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)
```



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")
```