

# Class 10: Structural Bioinformatics Pt.2

AUTHOR  
Hyeseung (Frankie) Son PID: A16025601

## Comparative Analysis of ADK

Adenylate kinase (Adk) is a ubiquitous enzyme that functions to maintain the equilibrium between cytoplasmic nucleotides essential for many cellular processes.

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

We will begin by getting an example ADK sequence from the database.

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?:

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

TRUE

library(bio3d)

Warning: package 'bio3d' was built under R version 4.2.3

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  DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDR
      61      .      .      .      .      .      120
```

```

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

      181      .      .      .      214
pdb|1AKE|A  YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG
      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.

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

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

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

```
#hits$pdb.id

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

```

Now we can download all these PDB strcuture files.

```
# 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 exists. Skipping download
```

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

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

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

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

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

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

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

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

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

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

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

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

	0%
=====	8%
=====	15%
=====	23%
=====	31%
=====	38%
=====	46%
=====	54%
=====	62%
=====	69%

```

|
|=====| 77%
|
|=====| 85%
|
|=====| 92%
|
|=====| 100%

```

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

```

# Align related PDBs
pdbbs <- pbdaln(files, fit = TRUE, exeFile="msa")

```

Reading PDB files:

```

pdbbs/split_chain/1AKE_A.pdb
pdbbs/split_chain/6S36_A.pdb
pdbbs/split_chain/6RZE_A.pdb
pdbbs/split_chain/3HPR_A.pdb
pdbbs/split_chain/1E4V_A.pdb
pdbbs/split_chain/5EJE_A.pdb
pdbbs/split_chain/1E4Y_A.pdb
pdbbs/split_chain/3X2S_A.pdb
pdbbs/split_chain/6HAP_A.pdb
pdbbs/split_chain/6HAM_A.pdb
pdbbs/split_chain/4K46_A.pdb
pdbbs/split_chain/3GMT_A.pdb
pdbbs/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: pdbbs/split_chain/1AKE_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 2  name: pdbbs/split_chain/6S36_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 3  name: pdbbs/split_chain/6RZE_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 4  name: pdbbs/split_chain/3HPR_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 5  name: pdbbs/split_chain/1E4V_A.pdb
pdb/seq: 6  name: pdbbs/split_chain/5EJE_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 7  name: pdbbs/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
```

```
[Truncated_Name:1]1AKE_A.pdb      1          .          .          .          40
-----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     TENLYFQSNMRIILLGAPGAGKGTQAKIIEQKYNIAHIS
                                *****  *****  *  *^ *  **
                                1          .          .          .          40

[Truncated_Name:1]1AKE_A.pdb      41          .          .          .          80
TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE
[Truncated_Name:2]6S36_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE
[Truncated_Name:3]6RZE_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE
[Truncated_Name:4]3HPR_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE
[Truncated_Name:5]1E4V_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE
[Truncated_Name:6]5EJE_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDACKLVTDDELVIALVKE
[Truncated_Name:7]1E4Y_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE
[Truncated_Name:8]3X2S_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDCGKLVTDDELVIALVKE
[Truncated_Name:9]6HAP_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVRE
[Truncated_Name:10]6HAM_A.pdb     TGDMLRAAIKSGSELGKQAKDIMDAGKLVTDDEIIIALVKE
[Truncated_Name:11]4K46_A.pdb     TGDMLRAAIKAGTELGKQAKSVIDAGQLVSDDIILGLVKE
[Truncated_Name:12]3GMT_A.pdb     TGDMLRAAVKAGTPLGVEAKTYMDEGKLVDPDSLIIGLVKE
[Truncated_Name:13]4PZL_A.pdb     TGDMIRETIKSGSALGQELKKVLDAGELVSDEFIIVKIVKD
***** ^* ^* **  *  ^*  ** *  ^^ ^^^^
                                41          .          .          .          80

[Truncated_Name:1]1AKE_A.pdb      81          .          .          .          120
RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:2]6S36_A.pdb      RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:3]6RZE_A.pdb      RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
```

```

[Truncated_Name:4]3HPR_A.pdb      RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:5]1E4V_A.pdb      RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:6]5EJE_A.pdb      RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:7]1E4Y_A.pdb      RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:8]3X2S_A.pdb      RIAQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:9]6HAP_A.pdb      RICQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:10]6HAM_A.pdb     RICQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:11]4K46_A.pdb     RIAQDDCAKGFLLDGFPRTIPQADGLKEVGVVVDYVIEFD
[Truncated_Name:12]3GMT_A.pdb     RLKEADCANGYLFDFGFPRTIAQADAMKEAGVAIDYVLEID
[Truncated_Name:13]4PZL_A.pdb     RISKNCNNGFLLDGVPRTIPQAQELDKLGVNIDYIVEVD
                                   *^  *   *^^ ** ***** ** ^   *^ ***** *
81                               .           .           120

121                             .           .           160

[Truncated_Name:1]1AKE_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:2]6S36_A.pdb      VPDELIVDKIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:3]6RZE_A.pdb      VPDELIVDAIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:4]3HPR_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDGTG
[Truncated_Name:5]1E4V_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:6]5EJE_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:7]1E4Y_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:8]3X2S_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:9]6HAP_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:10]6HAM_A.pdb     VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:11]4K46_A.pdb     VADSVIVERMAGRRRAHLASGRTYHNVPKVEGKDDVTG
[Truncated_Name:12]3GMT_A.pdb     VPFSEIIERMSGRRTHPASGRTYHVKNPPKVEGKDDVTG
[Truncated_Name:13]4PZL_A.pdb     VADNLLIERITGRIHPASGRTYHTKFNPPKVADKDDVTG
                                   *   ^^ ^ *** *   *** ** ^***** *** **
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      EELTTRKDDQEECVKRLVEYHQM 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 TAPLIYYGKEAEAGN
[Truncated_Name:12]3GMT_A.pdb     EPLVQRDDDKKEETVKKRLDVYEAQTKPLITYYGDWARRGA
[Truncated_Name:13]4PZL_A.pdb     EPLITRTDDNEDTVKQRLSVYHAQTAKLIDFYRNFSSNT
                                   * * * ** ^* ** *   *   ** ^*
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--QYLKFDGTPVAEVSAAELEKALA-
[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)
```

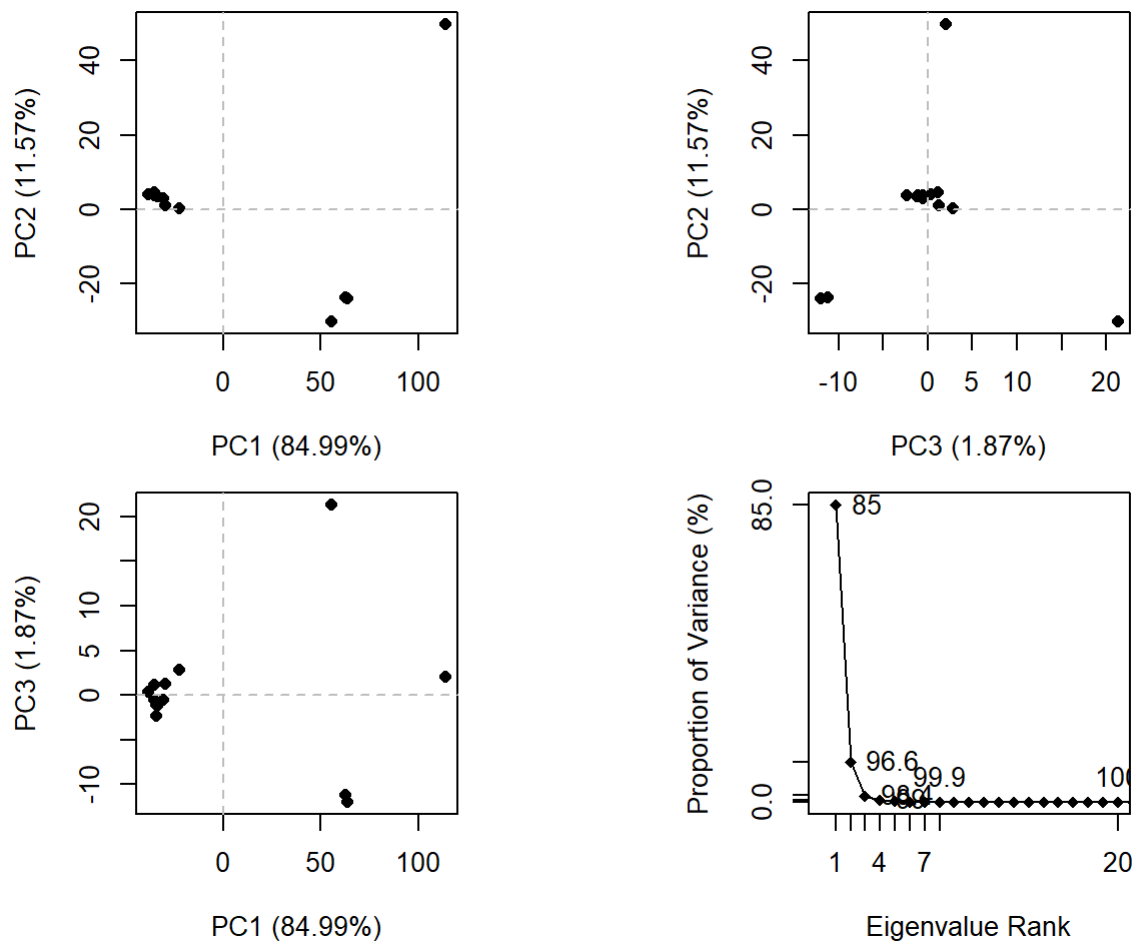
```
+ attr: xyz, resno, b, chain, id, ali, resid, sse, call
```

Now we have our aligned and superposed structures we can perform all sorts of analysis on them. 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 (Root Mean Square Distance) (or any other method).

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

Warning in rmsd(pdb): No indices provided, using the 204 non NA positions

```
rd
```

	1AKE_A	6S36_A	6RZE_A	3HPR_A	1E4V_A	5EJE_A	1E4Y_A	3X2S_A	6HAP_A	6HAM_A
1AKE_A	0.000	7.097	7.200	0.311	0.251	0.427	0.941	0.621	1.355	0.851
6S36_A	7.097	0.000	0.434	7.184	7.077	7.111	6.785	7.195	6.346	6.859
6RZE_A	7.200	0.434	0.000	7.289	7.178	7.212	6.882	7.290	6.441	6.959
3HPR_A	0.311	7.184	7.289	0.000	0.382	0.507	1.002	0.643	1.426	0.909
1E4V_A	0.251	7.077	7.178	0.382	0.000	0.475	0.971	0.636	1.377	0.851
5EJE_A	0.427	7.111	7.212	0.507	0.475	0.000	1.093	0.702	1.526	0.989
1E4Y_A	0.941	6.785	6.882	1.002	0.971	1.093	0.000	0.961	1.067	0.748
3X2S_A	0.621	7.195	7.290	0.643	0.636	0.702	0.961	0.000	1.448	0.899
6HAP_A	1.355	6.346	6.441	1.426	1.377	1.526	1.067	1.448	0.000	1.162
6HAM_A	0.851	6.859	6.959	0.909	0.851	0.989	0.748	0.899	1.162	0.000
4K46_A	1.013	7.420	7.518	0.951	1.039	1.059	1.182	0.870	1.787	1.019
3GMT_A	10.863	6.336	6.308	10.950	10.852	10.882	10.632	10.902	10.207	10.646

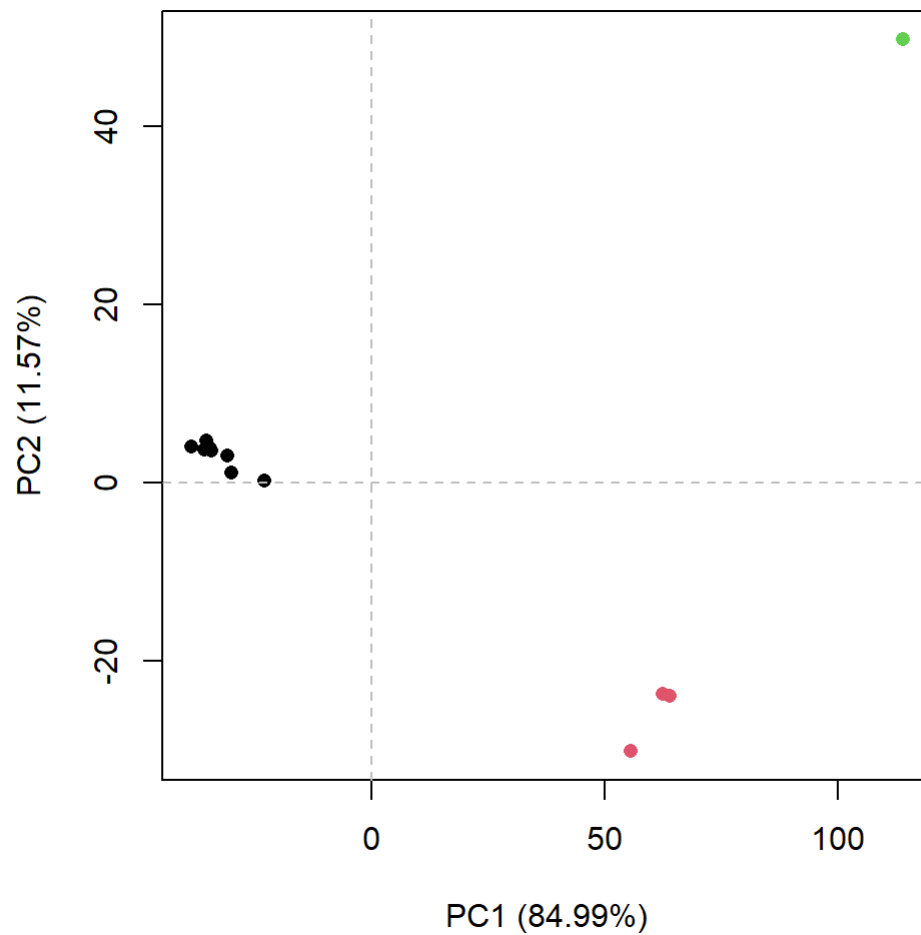


	4PZL_A	6.917	2.381	2.457	6.987	6.896	6.953	6.562	6.967	6.110	6.684
4K46_A	3GMT_A	4PZL_A									
1AKE_A	1.013	10.863	6.917								
6S36_A	7.420	6.336	2.381								
6RZE_A	7.518	6.308	2.457								
3HPR_A	0.951	10.950	6.987								
1E4V_A	1.039	10.852	6.896								
5EJE_A	1.059	10.882	6.953								
1E4Y_A	1.182	10.632	6.562								
3X2S_A	0.870	10.902	6.967								
6HAP_A	1.787	10.207	6.110								
6HAM_A	1.019	10.646	6.684								
4K46_A	0.000	11.156	7.199								
3GMT_A	11.156	0.000	7.047								
4PZL_A	7.199	7.047	0.000								

```
# Structure-based clustering
hc.rd <- hclust(dist(rd))

grps <- cutree(hc.rd, k=3)

plot(pc.xray, 1:2, col=grps, bg=grps)
```



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

```
# Visualize the Principal Component analysis  
pc1 <- mktroj(pc.xray, pc=1, file="pc_1.pdb")  
pc1
```

Total Frames#: 34

Total XYZs#: 612, (Atoms#: 204)

[1] 26.787 52.261 40.414 <...> 15.653 53.622 42.018 [20808]

+ attr: Matrix DIM = 34 x 612