Class 10: Structural Bioinformatics II

Raidah Anisah Huda

#Comparative analysis of ADK

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

There has been lots of work done on this protein due to it's importance including lots of crystal structures.

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

> 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?

We will begin with getting an example ADK seqeunce from the database.

```r
library(bio3d)
aa <- get.seq("1ake_A")
```

```
Warning in get.seq("1ake_A"): Removing existing file: seqs.fasta
```

```
Fetching... Please wait. Done.
```

```r
aa
```

```
             1        .         .         .         .         .         60
pdb|1AKE|A   MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
             1        .         .         .         .         .         60

            61        .         .         .         .         .         120
pdb|1AKE|A   DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRI
```

```
            61        .         .         .         .         .         120

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

We can run `hits`

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

Lets wee whats 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','
```

Now we can download all these PDB structure 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):
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
```

```
  |
  |                                                                      |   0%
  |
  |========                                                              |  11%
  |
  |================                                                      |  22%
  |
  |=======================                                               |  33%
  |
  |===============================                                       |  44%
  |
  |=======================================                               |  56%
  |
  |===============================================                       |  67%
  |
  |======================================================                |  78%
  |
  |==============================================================        |  89%
  |
  |======================================================================| 100%
```

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

```
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
   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
```

Let's have a look at how our `pdbs` object:

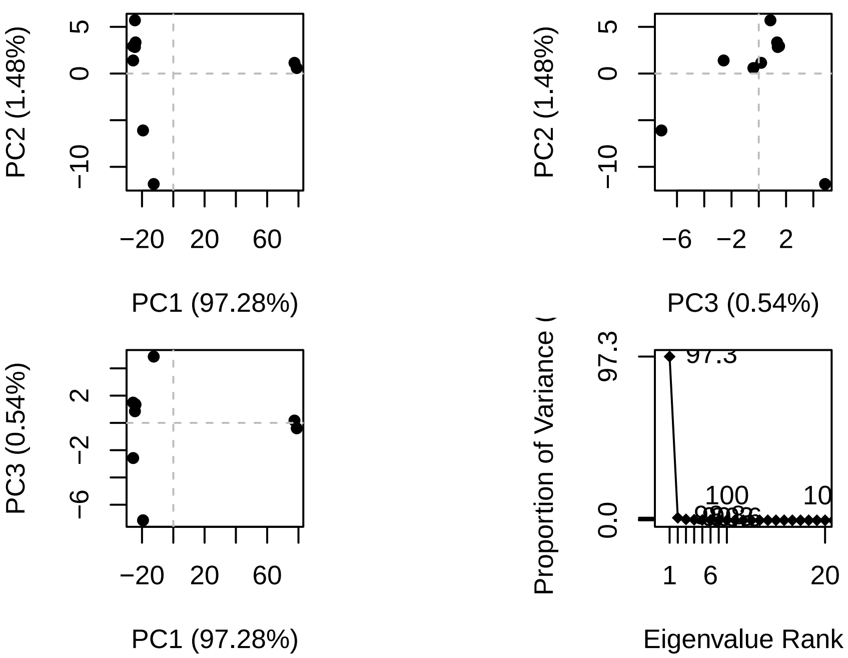
```
pdbs
```

```
                               1        .         .         .         .         50 
[Truncated_Name:1]1AKE_A.pdb   MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAK
[Truncated_Name:2]6S36_A.pdb   MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAK
[Truncated_Name:3]6RZE_A.pdb   MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAK
[Truncated_Name:4]3HPR_A.pdb   MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAK
[Truncated_Name:5]1E4V_A.pdb   MRIILLGAPVAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAK
[Truncated_Name:6]5EJE_A.pdb   MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAK
[Truncated_Name:7]1E4Y_A.pdb   MRIILLGALVAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAK
[Truncated_Name:8]3X2S_A.pdb   MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAK
[Truncated_Name:9]6HAP_A.pdb   MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAK
                               ********  **************************************** 
                               1        .         .         .         .         50 

                              51        .         .         .         .         100 
[Truncated_Name:1]1AKE_A.pdb   DIMDAGKLVTDELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAG
[Truncated_Name:2]6S36_A.pdb   DIMDAGKLVTDELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAG
[Truncated_Name:3]6RZE_A.pdb   DIMDAGKLVTDELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAG
[Truncated_Name:4]3HPR_A.pdb   DIMDAGKLVTDELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAG
[Truncated_Name:5]1E4V_A.pdb   DIMDAGKLVTDELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAG
[Truncated_Name:6]5EJE_A.pdb   DIMDACKLVTDELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAG
[Truncated_Name:7]1E4Y_A.pdb   DIMDAGKLVTDELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAG
[Truncated_Name:8]3X2S_A.pdb   DIMDCGKLVTDELVIALVKERIAQEDSRNGFLLDGFPRTIPQADAMKEAG
[Truncated_Name:9]6HAP_A.pdb   DIMDAGKLVTDELVIALVRERICQEDSRNGFLLDGFPRTIPQADAMKEAG
                               ****  ************^*** *** *********************** 
                              51        .         .         .         .         100 

                             101        .         .         .         .         150 
[Truncated_Name:1]1AKE_A.pdb   INVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:2]6S36_A.pdb   INVDYVLEFDVPDELIVDKIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:3]6RZE_A.pdb   INVDYVLEFDVPDELIVDAIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:4]3HPR_A.pdb   INVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDGTG
[Truncated_Name:5]1E4V_A.pdb   INVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:6]5EJE_A.pdb   INVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:7]1E4Y_A.pdb   INVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:8]3X2S_A.pdb   INVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:9]6HAP_A.pdb   INVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
```

```
                               ****************** **************************** **
                             101        .         .         .         .         150

                             151        .         .         .         .         200
[Truncated_Name:1]1AKE_A.pdb   EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDGTK
[Truncated_Name:2]6S36_A.pdb   EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDGTK
[Truncated_Name:3]6RZE_A.pdb   EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDGTK
[Truncated_Name:4]3HPR_A.pdb   EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDGTK
[Truncated_Name:5]1E4V_A.pdb   EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDGTK
[Truncated_Name:6]5EJE_A.pdb   EELTTRKDDQEECVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDGTK
[Truncated_Name:7]1E4Y_A.pdb   EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDGTK
[Truncated_Name:8]3X2S_A.pdb   EELTTRKDDQEETVRKRLCEYHQMTAPLIGYYSKEAEAGNTKYAKVDGTK
[Truncated_Name:9]6HAP_A.pdb   EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDGTK
                               ************ ***** *******************************
                             151        .         .         .         .         200

                             201        .   214
[Truncated_Name:1]1AKE_A.pdb   PVAEVRADLEKILG
[Truncated_Name:2]6S36_A.pdb   PVAEVRADLEKILG
[Truncated_Name:3]6RZE_A.pdb   PVAEVRADLEKILG
[Truncated_Name:4]3HPR_A.pdb   PVAEVRADLEKILG
[Truncated_Name:5]1E4V_A.pdb   PVAEVRADLEKILG
[Truncated_Name:6]5EJE_A.pdb   PVAEVRADLEKILG
[Truncated_Name:7]1E4Y_A.pdb   PVAEVRADLEKILG
[Truncated_Name:8]3X2S_A.pdb   PVAEVRADLEKILG
[Truncated_Name:9]6HAP_A.pdb   PVCEVRADLEKILG
                               ** ***********
                             201        .   214

Call:
  pdbaln(files = files, fit = TRUE, exefile = "msa")

Class:
  pdbs, fasta

Alignment dimensions:
  9 sequence rows; 214 position columns (214 non-gap, 0 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(pdbs)
plot(pc.xray)
```



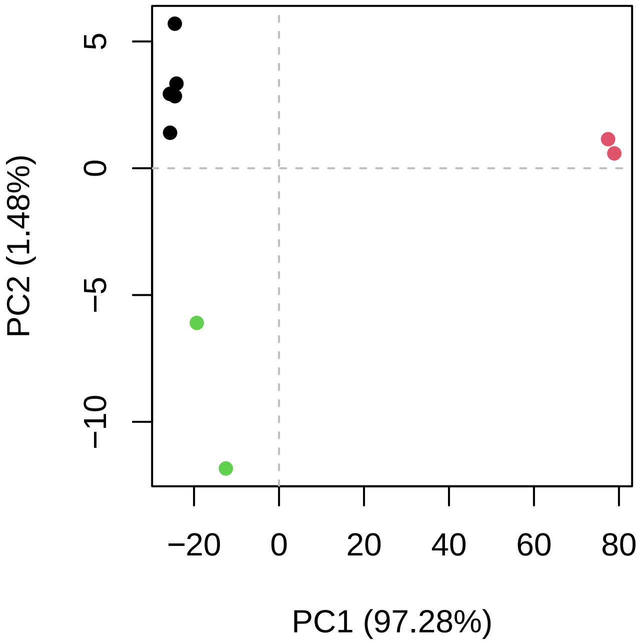
Presults of PCA an Adenlyat kinase X-ray structures. Each dot represents one PDB structures,

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

```
rd <- rmsd(pdbs)
```

```
Warning in rmsd(pdbs): No indices provided, using the 214 non NA positions
```

```
# Structure-based clustering
hc.rd <- hclust(dist(rd))
grps.rd <- cutree(hc.rd, k=3)

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



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

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