Class 10 (pt 2)

Brian Wells (PID: A69026838)

We'll finish off lab 10

##Comparative structure analysis of Adenylate Kinase

We will use `bio3d` package for this analysis that starts with a single sequence . We will also use the `msa` package from BioConductor. First we need to install the `BiocManager` we install it same as we always do (`install.packages()`)

We use `BiocManager::install()` to install any other BioConductor package we want - like `msa` in this case.

```
library(bio3d)
```

First we will pick a sequence "1ake_A" and use `get.seq()`

```
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   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
```

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

214

I want to seqarch for all related structures

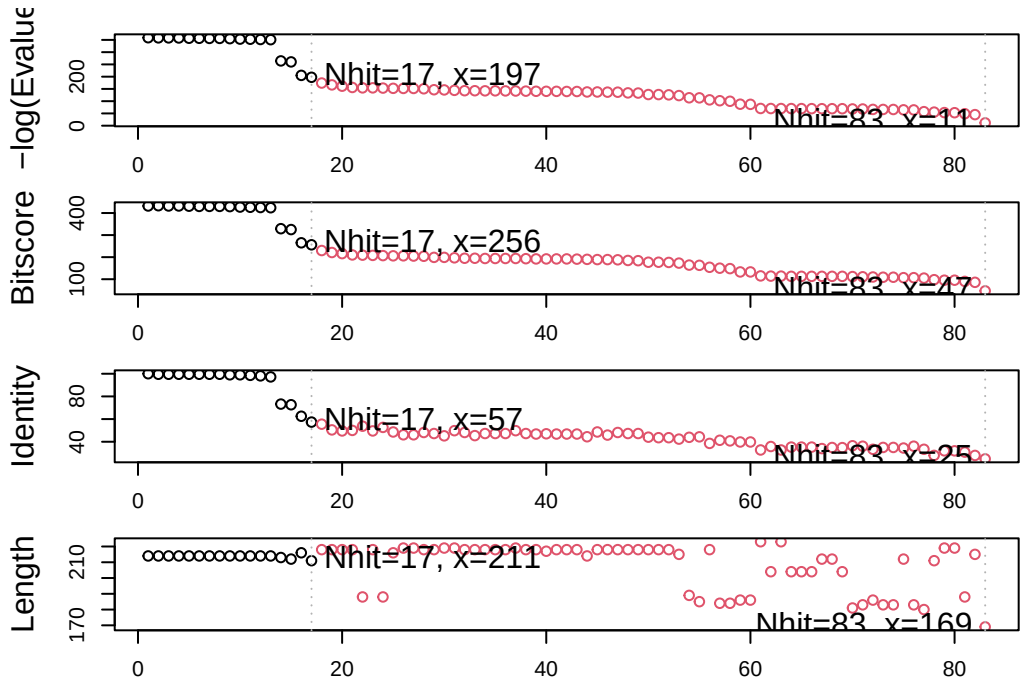
```
b <- blast.pdb(aa)
```

```
Searching ... please wait (updates every 5 seconds) RID = MS6FVJG1013
.
Reporting 83 hits
```

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

```
 * Possible cutoff values:    197 11
           Yielding Nhits:    17 83

 * Chosen cutoff value of:    197
           Yielding Nhits:    17
```




```
head(b$hit.tbl)
```

```
     queryid subjectids identity alignmentlength mismatches gapopens q.start
1 Query_5753     1AKE_A  100.000             214          0        0       1
2 Query_5753     8BQF_A   99.533             214          1        0       1
3 Query_5753     4X8M_A   99.533             214          1        0       1
4 Query_5753     6S36_A   99.533             214          1        0       1
5 Query_5753     6RZE_A   99.533             214          1        0       1
6 Query_5753     4X8H_A   99.533             214          1        0       1
  q.end s.start s.end    evalue bitscore positives mlog.evalue pdb.id    acc
1   214       1   214 1.45e-156      432    100.00    358.8317 1AKE_A 1AKE_A
2   214      21   234 2.38e-156      433    100.00    358.3362 8BQF_A 8BQF_A
3   214       1   214 2.60e-156      432    100.00    358.2478 4X8M_A 4X8M_A
4   214       1   214 3.82e-156      432    100.00    357.8630 6S36_A 6S36_A
5   214       1   214 1.10e-155      431     99.53    356.8054 6RZE_A 6RZE_A
6   214       1   214 1.44e-155      430     99.53    356.5360 4X8H_A 4X8H_A
```

```
hits$pdb.id
```

```
[1] "1AKE_A" "8BQF_A" "4X8M_A" "6S36_A" "6RZE_A" "4X8H_A" "3HPR_A" "1E4V_A"
```

```
 [9] "5EJE_A" "1E4Y_A" "3X2S_A" "6HAP_A" "6HAM_A" "4K46_A" "4NP6_A" "3GMT_A"
[17] "4PZL_A"
```

```
save(hits, b, file="blast_results.Rds")
```

```
load("blast_results.Rds")
```

Now we will download all the related structures from the database with `getpds()`

Search and retrieve ADK structures

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

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

```
Warning in get.pdb(hits$pdb.id, path = "pdbs", split = T, gzip = T):
pdbs/1AKE.pdb.gz exists. Skipping download
```

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

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

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

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

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

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

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

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

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

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

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

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


  |
  |                                                                      |   0%
  |
  |=====                                                                 |   8%
  |
  |===========                                                           |  15%
  |
  |================                                                      |  23%
  |
  |======================                                                |  31%
  |
  |============================                                          |  38%
  |
  |================================                                      |  46%
  |
  |====================================                                  |  54%
  |
  |============================================                          |  62%
  |
  |================================================                      |  69%
  |
  |========================================================              |  77%
```

```
  |
  |=============================================================           |  85%
  |
  |=================================================================       |  92%
  |
  |=====================================================================| 100%
```

#Align and superpose structures

```
# Align releated PDBs
pdbs <- pdbaln(files, fit = T, 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
```

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

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

#Annotate collected PDB structures

```
anno <- pdb.annotate(ids)
unique(anno$source)
```

```
[1] "Escherichia coli"
[2] "Escherichia coli K-12"
[3] "Escherichia coli O139:H28 str. E24377A"
[4] "Escherichia coli str. K-12 substr. MDS42"
[5] "Photobacterium profundum"
[6] "Burkholderia pseudomallei 1710b"
[7] "Francisella tularensis subsp. tularensis SCHU S4"
```

```
anno
```

```
       structureId chainId macromoleculeType chainLength experimentalTechnique
1AKE_A        1AKE       A           Protein         214                 X-ray
6S36_A        6S36       A           Protein         214                 X-ray
```

6RZE_A	6RZE	A	Protein	214		X-ray
3HPR_A	3HPR	A	Protein	214		X-ray
1E4V_A	1E4V	A	Protein	214		X-ray
5EJE_A	5EJE	A	Protein	214		X-ray
1E4Y_A	1E4Y	A	Protein	214		X-ray
3X2S_A	3X2S	A	Protein	214		X-ray
6HAP_A	6HAP	A	Protein	214		X-ray
6HAM_A	6HAM	A	Protein	214		X-ray
4K46_A	4K46	A	Protein	214		X-ray
3GMT_A	3GMT	A	Protein	230		X-ray
4PZL_A	4PZL	A	Protein	242		X-ray

	resolution	scopDomain	pfam
1AKE_A	2.00	Adenylate kinase	Adenylate kinase, active site lid (ADK_lid)
6S36_A	1.60	<NA>	Adenylate kinase, active site lid (ADK_lid)
6RZE_A	1.69	<NA>	Adenylate kinase, active site lid (ADK_lid)
3HPR_A	2.00	<NA>	Adenylate kinase, active site lid (ADK_lid)
1E4V_A	1.85	Adenylate kinase	Adenylate kinase, active site lid (ADK_lid)
5EJE_A	1.90	<NA>	Adenylate kinase, active site lid (ADK_lid)
1E4Y_A	1.85	Adenylate kinase	Adenylate kinase, active site lid (ADK_lid)
3X2S_A	2.80	<NA>	Adenylate kinase, active site lid (ADK_lid)
6HAP_A	2.70	<NA>	Adenylate kinase, active site lid (ADK_lid)
6HAM_A	2.55	<NA>	Adenylate kinase, active site lid (ADK_lid)
4K46_A	2.01	<NA>	Adenylate kinase, active site lid (ADK_lid)
3GMT_A	2.10	<NA>	Adenylate kinase, active site lid (ADK_lid)
4PZL_A	2.10	<NA>	Adenylate kinase, active site lid (ADK_lid)

	ligandId	ligandName
1AKE_A	AP5	BIS(ADENOSINE)-5'-PENTAPHOSPHATE
6S36_A	CL (3),NA,MG (2)	CHLORIDE ION (3),SODIUM ION,MAGNESIUM ION (2)
6RZE_A	NA (3),CL (2)	SODIUM ION (3),CHLORIDE ION (2)
3HPR_A	AP5	
1E4V_A	AP5	
5EJE_A	AP5,CO	
1E4Y_A	AP5	
3X2S_A	JPY (2),AP5,MG	
6HAP_A	AP5	
6HAM_A	AP5	
4K46_A	ADP,AMP,PO4	
3GMT_A	SO4 (2)	
4PZL_A	CA,FMT,GOL	

3HPR_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE
1E4V_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE
5EJE_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE,COBALT (II) ION
1E4Y_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE
3X2S_A	N-(pyren-1-ylmethyl)acetamide (2),BIS(ADENOSINE)-5'-PENTAPHOSPHATE,MAGNESIUM ION
6HAP_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE
6HAM_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE
4K46_A	ADENOSINE-5'-DIPHOSPHATE,ADENOSINE MONOPHOSPHATE,PHOSPHATE ION
3GMT_A	SULFATE ION (2)
4PZL_A	CALCIUM ION,FORMIC ACID,GLYCEROL

	source
1AKE_A	Escherichia coli
6S36_A	Escherichia coli
6RZE_A	Escherichia coli
3HPR_A	Escherichia coli K-12
1E4V_A	Escherichia coli
5EJE_A	Escherichia coli O139:H28 str. E24377A
1E4Y_A	Escherichia coli
3X2S_A	Escherichia coli str. K-12 substr. MDS42
6HAP_A	Escherichia coli O139:H28 str. E24377A
6HAM_A	Escherichia coli K-12
4K46_A	Photobacterium profundum
3GMT_A	Burkholderia pseudomallei 1710b
4PZL_A	Francisella tularensis subsp. tularensis SCHU S4

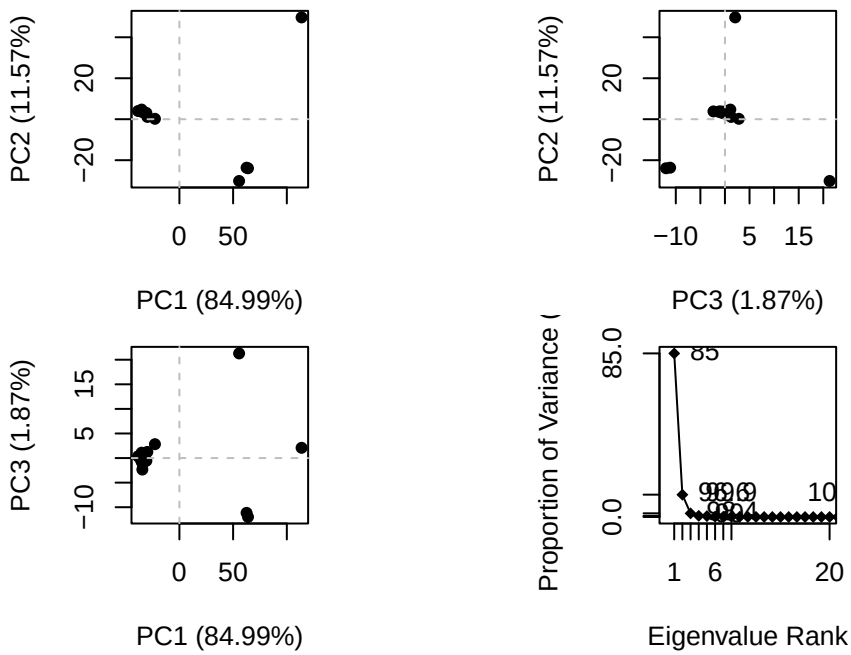
1AKE_A	STRUCTURE OF THE COMPLEX BETWEEN ADENYLATE KINASE FROM ESCHERICHIA COLI AND THE INHIBI
6S36_A	
6RZE_A	
3HPR_A	
1E4V_A	
5EJE_A	Cryst
1E4Y_A	
3X2S_A	
6HAP_A	
6HAM_A	
4K46_A	
3GMT_A	
4PZL_A	The cryst

	citation	rObserved	rFree
1AKE_A	Muller, C.W., et al. J Mol Biol (1992)	0.19600	NA
6S36_A	Rogne, P., et al. Biochemistry (2019)	0.16320	0.23560
6RZE_A	Rogne, P., et al. Biochemistry (2019)	0.18650	0.23500
3HPR_A	Schrank, T.P., et al. Proc Natl Acad Sci U S A (2009)	0.21000	0.24320

```
1E4V_A                    Muller, C.W., et al. Proteins (1993)   0.19600      NA
5EJE_A  Kovermann, M., et al. Proc Natl Acad Sci U S A (2017)   0.18890 0.23580
1E4Y_A                    Muller, C.W., et al. Proteins (1993)   0.17800      NA
3X2S_A                 Fujii, A., et al. Bioconjug Chem (2015)   0.20700 0.25600
6HAP_A                Kantaev, R., et al. J Phys Chem B (2018)   0.22630 0.27760
6HAM_A                Kantaev, R., et al. J Phys Chem B (2018)   0.20511 0.24325
4K46_A                     Cho, Y.-J., et al. To be published    0.17000 0.22290
3GMT_A Buchko, G.W., et al. Biochem Biophys Res Commun (2010)   0.23800 0.29500
4PZL_A                       Tan, K., et al. To be published    0.19360 0.23680
         rWork spaceGroup
1AKE_A 0.19600  P 21 2 21
6S36_A 0.15940    C 1 2 1
6RZE_A 0.18190    C 1 2 1
3HPR_A 0.20620  P 21 21 2
1E4V_A 0.19600  P 21 2 21
5EJE_A 0.18630  P 21 2 21
1E4Y_A 0.17800   P 1 21 1
3X2S_A 0.20700 P 21 21 21
6HAP_A 0.22370    I 2 2 2
6HAM_A 0.20311       P 43
4K46_A 0.16730 P 21 21 21
3GMT_A 0.23500   P 1 21 1
4PZL_A 0.19130       P 32
```

#Principal component analysis

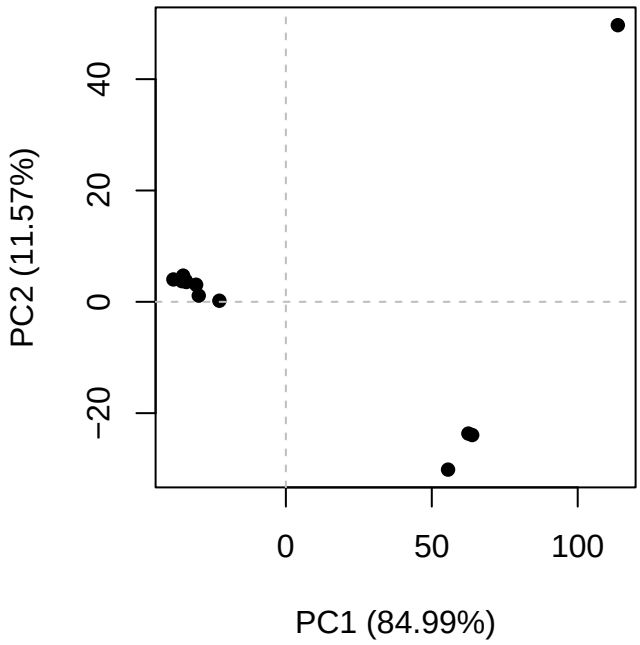
```
# Perform PCA
pc.xray <- pca(pdbs)
plot(pc.xray)
```




```
dim(pdbs$xyz)
```

```
[1]  13 681
```

```
plot(pc.xray, 1:2)
```

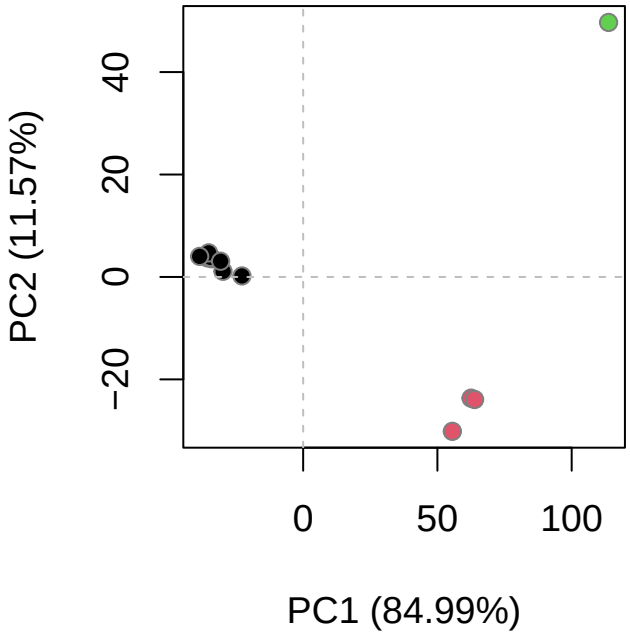



```
# Calculate RMSD
rd <- rmsd(pdbs)
```

```
Warning in rmsd(pdbs): 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)
```



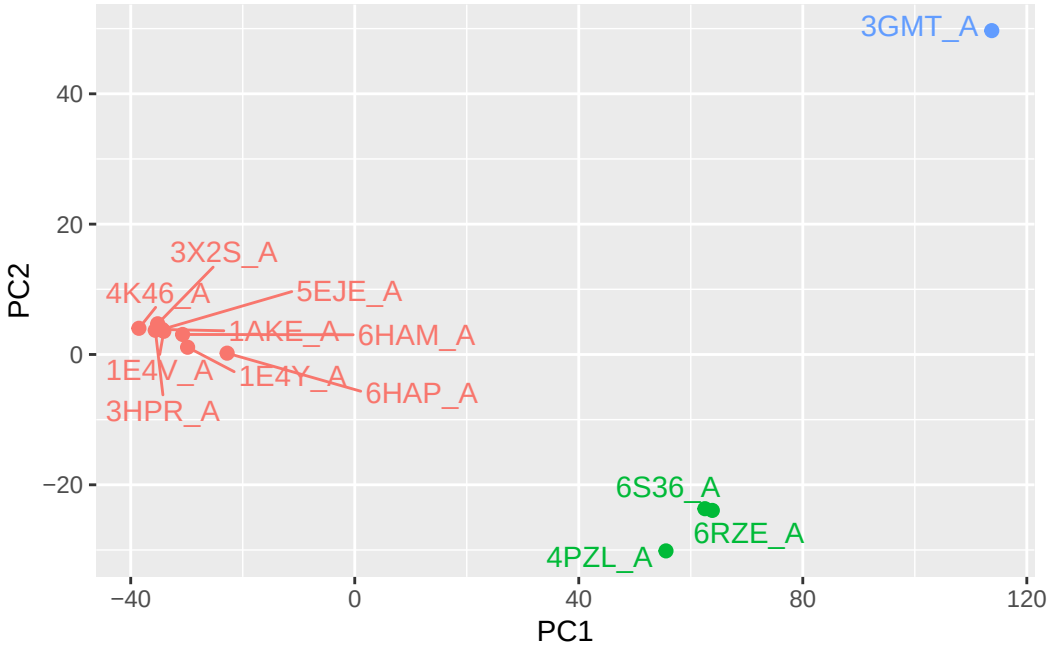
#Optional: further viewing optimization

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

```
#Plotting results with ggplot2
library(ggplot2)
library(ggrepel)

df <- data.frame(PC1=pc.xray$z[,1],
                 PC2=pc.xray$z[,2],
                 col=as.factor(grps.rd),
                 ids=ids)

p <- ggplot(df) +
  aes(PC1, PC2, col=col, label=ids) +
  geom_point(size=2) +
  geom_text_repel(max.overlaps = 20) +
  theme(legend.position = "none")
p
```



#Optional: Normal Analysis Mode

```
# NMA of all structures
modes <- nma(pdbs)
```

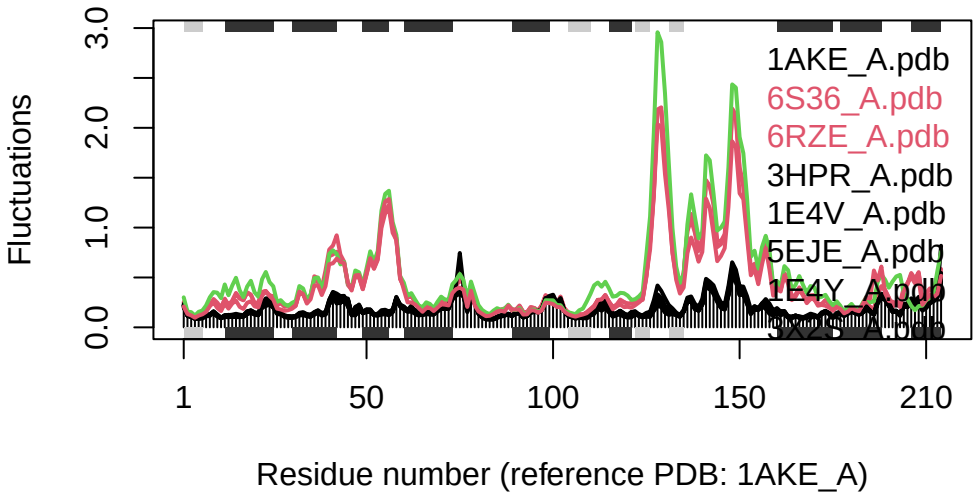
```
Details of Scheduled Calculation:
  ... 13 input structures
  ... storing 606 eigenvectors for each structure
  ... dimension of x$U.subspace: ( 612x606x13 )
  ... coordinate superposition prior to NM calculation
  ... aligned eigenvectors (gap containing positions removed)
  ... estimated memory usage of final 'eNMA' object: 36.9 Mb


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

```
  |================                                                      |  23%
  |                                                                      |
  |=====================                                                 |  31%
  |                                                                      |
  |==========================                                            |  38%
  |                                                                      |
  |================================                                      |  46%
  |                                                                      |
  |=======================================                               |  54%
  |                                                                      |
  |==============================================                        |  62%
  |                                                                      |
  |=====================================================                 |  69%
  |                                                                      |
  |============================================================          |  77%
  |                                                                      |
  |===================================================================   |  85%
  |                                                                      |
  |=========================================================================     |  92%
  |                                                                      |
  |================================================================================| 100%
```

```
plot(modes, pdbs, col=grps.rd)
```

```
Extracting SSE from pdbs$sse attribute
```



Q14. What do you note about this plot? Are the black and colored lines similar or different? Where do you think they differ most and why?

Generally the peaks at certain residue numbers line up, but the magnitude (mainly between the black and colored lines) differ greatly. They differ most between residue ~125 - ~150. This could be t