

# Class 11: Comparative Structure Analysis

Emily Chen (PID: A16925878)

Load up the packages we will need for the analysis of protein structure sets.

```
library(bio3d)
```

We will analyze the ADK starting with a single ADK database accession code: "lake\_A"

```
# id <- get.seq("lake_A")
id <- 'lake_A'
aa <- get.seq(id)
```

Warning in get.seq(id): Removing existing file: seqs.fasta

Fetching... Please wait. Done.

```
aa
```

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

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

Now we can search the PDB database to find all related entries

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

Searching ... please wait (updates every 5 seconds) RID = 1WH1N6NS016

.....  
Reporting 90 hits

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

	queryid	subjectids	identity	alignmentlength	mismatches	gapopens	q.start
1	Query_5025243	1AKE_A	100.000	214	0	0	1
2	Query_5025243	8BQF_A	99.533	214	1	0	1
3	Query_5025243	4X8M_A	99.533	214	1	0	1
4	Query_5025243	6S36_A	99.533	214	1	0	1
5	Query_5025243	8Q2B_A	99.533	214	1	0	1
6	Query_5025243	8RJ9_A	99.533	214	1	0	1

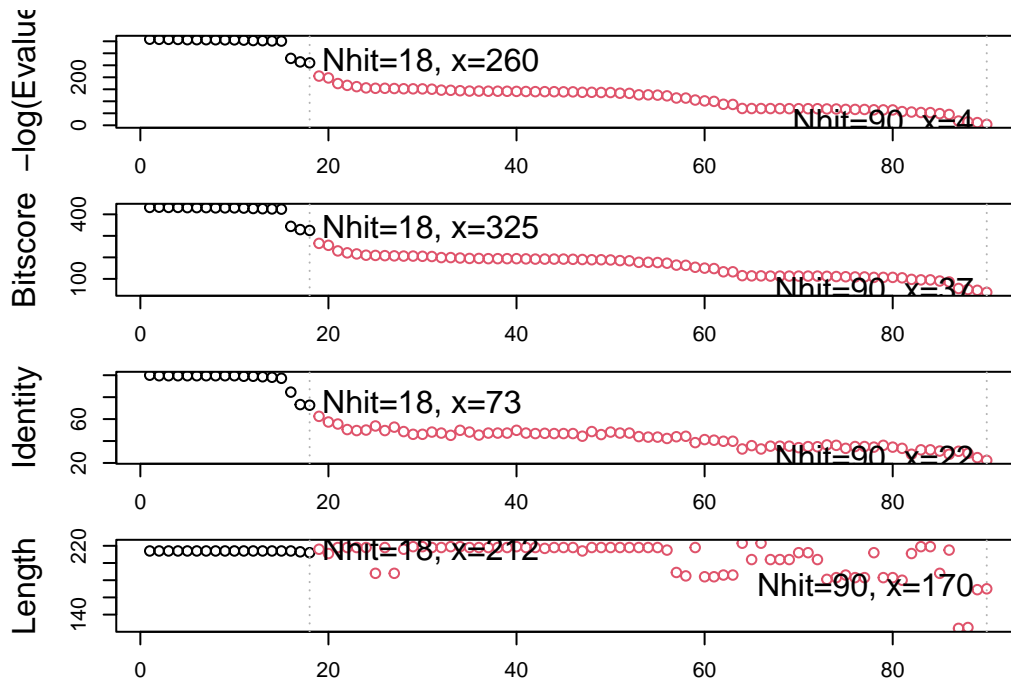
	q.end	s.start	s.end	evaluate	bitscore	positives	mlog.evaluate	pdb.id	acc
1	214	1	214	1.66e-156	432	100.00	358.6965	1AKE_A	1AKE_A
2	214	21	234	2.71e-156	433	100.00	358.2063	8BQF_A	8BQF_A
3	214	1	214	2.96e-156	432	100.00	358.1181	4X8M_A	4X8M_A
4	214	1	214	4.35e-156	432	100.00	357.7331	6S36_A	6S36_A
5	214	1	214	1.15e-155	431	99.53	356.7609	8Q2B_A	8Q2B_A
6	214	1	214	1.15e-155	431	99.53	356.7609	8RJ9_A	8RJ9_A

make a little summary figure of these results

```
hits <- plot(blast)
```

```
* Possible cutoff values: 260 3
    Yielding Nhits: 18 90
```

```
* Chosen cutoff value of: 260
    Yielding Nhits: 18
```



The “top hits” are our most similar entries in the resultst

```
hits$ pdb.id
```

```
[1] "1AKE_A" "8BQF_A" "4X8M_A" "6S36_A" "8Q2B_A" "8RJ9_A" "6RZE_A" "4X8H_A"
[9] "3HPR_A" "1E4V_A" "5EJE_A" "1E4Y_A" "3X2S_A" "6HAP_A" "6HAM_A" "8PVW_A"
[17] "4K46_A" "4NP6_A"
```

```
#download related PDB files
```

```
files <- get.pdb(hits$ pdb.id, path="pdbc", split=TRUE, gzip=TRUE)
```

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

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
pdbs/8BQF.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
pdbs/4X8M.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/8Q2B.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
pdbs/8RJ9.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/4X8H.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

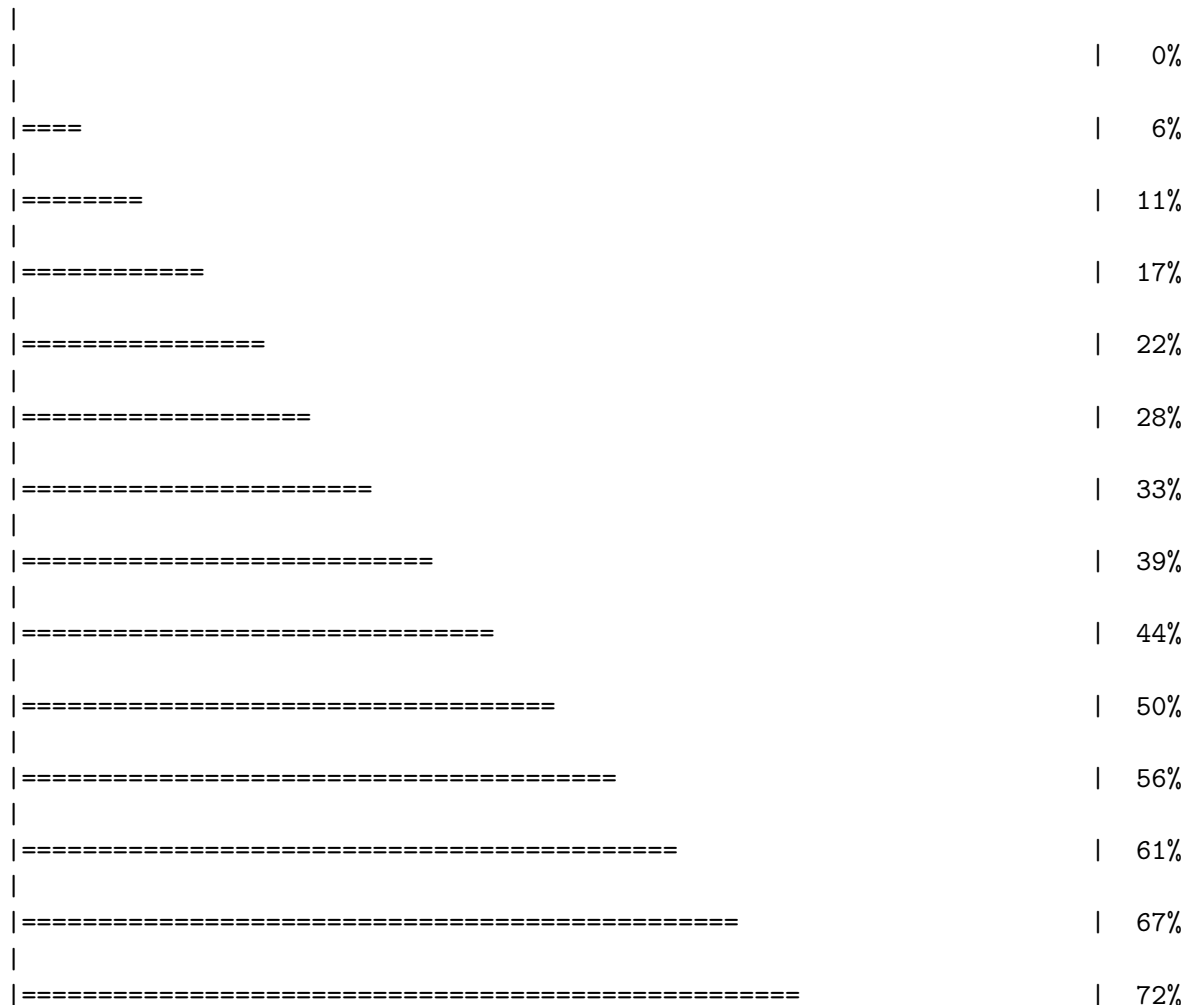
Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
pdbs/6HAP.pdb.gz exists. Skipping download

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/8PVW.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/4NP6.pdb.gz exists. Skipping download



```

|
|=====| 78%
|
|=====| 83%
|
|=====| 89%
|
|=====| 94%
|
|=====| 100%

```

```

# Align related PBDs
pdbbs <- pdbaln(files, fit = TRUE, exefile="msa")

```

Reading PDB files:

```

pdbbs/split_chain/1AKE_A.pdb
pdbbs/split_chain/8BQF_A.pdb
pdbbs/split_chain/4X8M_A.pdb
pdbbs/split_chain/6S36_A.pdb
pdbbs/split_chain/8Q2B_A.pdb
pdbbs/split_chain/8RJ9_A.pdb
pdbbs/split_chain/6RZE_A.pdb
pdbbs/split_chain/4X8H_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/8PVW_A.pdb
pdbbs/split_chain/4K46_A.pdb
pdbbs/split_chain/4NP6_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
..   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

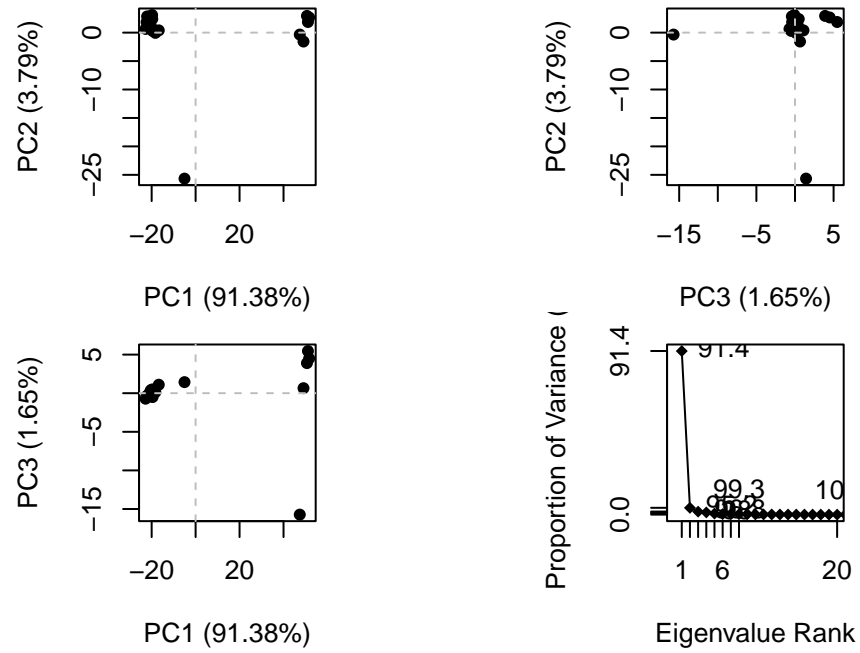
```
pdb/seq: 1   name: pdbs/split_chain/1AKE_A.pdb
            PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 2   name: pdbs/split_chain/8BQF_A.pdb
            PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 3   name: pdbs/split_chain/4X8M_A.pdb
pdb/seq: 4   name: pdbs/split_chain/6S36_A.pdb
            PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 5   name: pdbs/split_chain/8Q2B_A.pdb
            PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 6   name: pdbs/split_chain/8RJ9_A.pdb
            PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 7   name: pdbs/split_chain/6RZE_A.pdb
            PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 8   name: pdbs/split_chain/4X8H_A.pdb
pdb/seq: 9   name: pdbs/split_chain/3HPR_A.pdb
            PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 10  name: pdbs/split_chain/1E4V_A.pdb
pdb/seq: 11  name: pdbs/split_chain/5EJE_A.pdb
            PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 12  name: pdbs/split_chain/1E4Y_A.pdb
pdb/seq: 13  name: pdbs/split_chain/3X2S_A.pdb
pdb/seq: 14  name: pdbs/split_chain/6HAP_A.pdb
pdb/seq: 15  name: pdbs/split_chain/6HAM_A.pdb
            PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 16  name: pdbs/split_chain/8PVW_A.pdb
            PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 17  name: pdbs/split_chain/4K46_A.pdb
            PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 18  name: pdbs/split_chain/4NP6_A.pdb
```

```
library(bio3dview)
#view.pdbs(pdbs)
```

This view is better than all the data; however, it is still hard to see the alignment and to see which one is more similar to the other or how to families of sequences are similar.

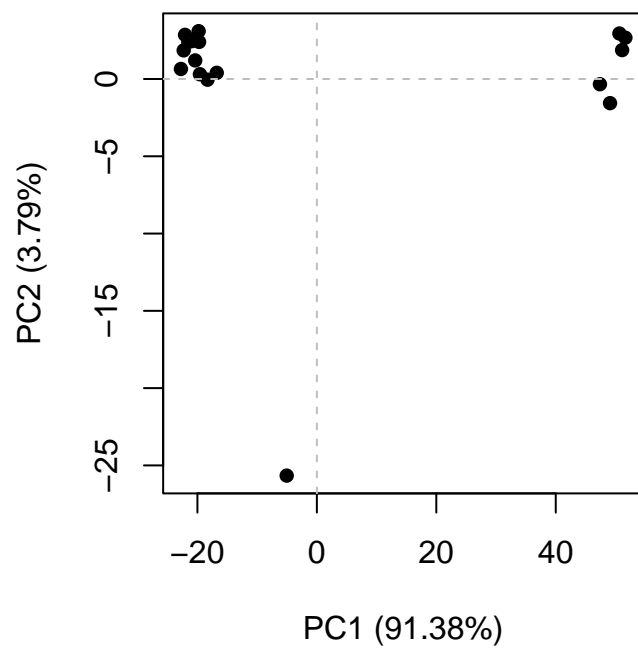
Lets do a PCA:

```
pc<- pca(pdfs)
plot(pc)
```



```
plot(pc,pc.axes=1:2)
```





Write a PDB “trajectory” for mol-star

```
mktrj(pc, fil="pca_results.pdb")
```