

Section 4 - Comparative Structural Analysis

Pt.1

Isabel Philip (A16855684)

Here we run through a complete “pipeline” of structural analysis that begins with a single sequence identifier and ends in a PCA analysis.

```
library(bio3d)
```

Step 1. Get a sequence

Step 1: Create a sequence for the protein we are interested in. We will take ADK “1ake_A”

```
id <- "1ake_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  DELVIALVKERIAQEDCRNGFLLDGFPRTPQADAMKEAGINVDYVLEFDVPDELIVDRI  
      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
```

Step 2. Run a Blast Search

Run a BLAST search of the PDB for all related sequences to our input **aa**

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

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

```
hits <- NULL
```

```
hits$ pdb.id <- c('1AKE_A', '6S36_A', '6RZE_A', '3HPR_A', '1E4V_A', '5EJE_A', '1E4Y_A', '3X2S_A', '6HAP_A', '6HAM_A', '4K46_A', '3GMT_A', '4PZL_A')
```

Step 3. Download Structures

These are the top hit (18) - i.e. all the structures in the PDB database related to our input sequence.

```
hits$ pdb.id
```

```
[1] "1AKE_A" "6S36_A" "6RZE_A" "3HPR_A" "1E4V_A" "5EJE_A" "1E4Y_A" "3X2S_A"
[9] "6HAP_A" "6HAM_A" "4K46_A" "3GMT_A" "4PZL_A"
```

```
# 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 = "pdb", split = TRUE, gzip = TRUE):
pdb/6S36.pdb.gz exists. Skipping download
```

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

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

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

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

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

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

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

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

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

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

	0%
=====	8%
=====	15%
=====	23%
=====	31%
=====	38%
=====	46%
=====	54%
=====	62%
=====	69%
=====	77%
=====	85%
=====	92%
=====	100%

Step 4. Align and Sequence

```
# 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

```

```

[Truncated_Name:1] 1AKE_A.pdb
[Truncated_Name:2] 6S36_A.pdb
[Truncated_Name:3] 6RZE_A.pdb
[Truncated_Name:4] 3HPR_A.pdb
[Truncated_Name:5] 1E4V_A.pdb
[Truncated_Name:6] 5EJE_A.pdb
[Truncated_Name:7] 1E4Y_A.pdb
[Truncated_Name:8] 3X2S_A.pdb
[Truncated_Name:9] 6HAP_A.pdb
[Truncated_Name:10] 6HAM_A.pdb
[Truncated_Name:11] 4K46_A.pdb
[Truncated_Name:12] 3GMT_A.pdb
[Truncated_Name:13] 4PZL_A.pdb

1          .          .          .          40
-----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
-----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
-----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
-----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
-----MRIILLGAPVAGKGTQAQFIMEKYGIPQIS
-----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
-----MRIILLGALVAGKGTQAQFIMEKYGIPQIS
-----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
-----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
-----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
-----MRIILLGAPGAGKGTQAQFIMAKFGIPQIS
-----MRLILLGAPGAGKGTQANFIKEKFGIPQIS
TENLYFQSNAMRIILLGAPGAGKGTQAKIIEQKYNIAHIS
          **~*****  *****  *  *~ *  **
1          .          .          .          40

41          .          .          .          80
TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
TGDMLRAAVKSGSELGKQAKDIMDACKLVDELVIALVKE
TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE
TGDMLRAAVKSGSELGKQAKDIMDCGKLVDELVIALVKE
TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVRE
TGDMLRAAIAKSGSELGKQAKDIMDAGKLVDEIIIALVKE
TGDMLRAAIAKAGTELGKQAKSVIDAGQLVSDDIILGLVKE
TGDMLRAAVKAGTPLGVEAKTYMDEGKLVPSLIIGLVKE
TGDMIRETIKSGSALGQELKKVLDAGELVSDEFIIVKVD
****~*  ~* *~ **  *  ~*  ** *  ^^ ~* ^^
41          .          .          .          80

81          .          .          .          120
RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
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 TIPQADGLKEVGVVVDYVIEFD
[Truncated_Name:12] 3GMT_A.pdb	RLKEADCANGYLF DGFPR TIAQADAMKEAGVAIDYVLEID
[Truncated_Name:13] 4PZL_A.pdb	RISKND CNNGFLLDGVPR TIPQAQELDKLGVNIDYIVEVD
	*^ * *~* ** ***** * ^ *~ ~**~* *
	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	VADSVIVERMAGRRAHLASGR TYHNVPKVEGKDDVTG
[Truncated_Name:12] 3GMT_A.pdb	VPFSEIIERMSGRRTHPASGR TYHV KFNPPKVEGKDDVTG
[Truncated_Name:13] 4PZL_A.pdb	VADNLLIERITGRRIHPASGR TYHTKFNPPKVADKDDVTG
	* ^^^ ^ *** * *** * ^***** *** *
	121 . . . 160
	161 . . . 200
[Truncated_Name:1] 1AKE_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:2] 6S36_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:3] 6RZE_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:4] 3HPR_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:5] 1E4V_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:6] 5EJE_A.pdb	EELTTRKDDQEECVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:7] 1E4Y_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:8] 3X2S_A.pdb	EELTTRKDDQEETVRKRLCEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:9] 6HAP_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:10] 6HAM_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:11] 4K46_A.pdb	EDLVIREDDKEETV LARLG VYHNQTAPLIAYYGKEAEAGN
[Truncated_Name:12] 3GMT_A.pdb	EPLVQRDDKEETVKKRLDVYEAQTKPLITYYGDWARRGA
[Truncated_Name:13] 4PZL_A.pdb	EPLITRTDDNEDTVKQRLSVYHAQTAKLIDFYRNFSSNT
	* * * * * ~ * * * * * ~ *

```

161          .          .          .          200

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

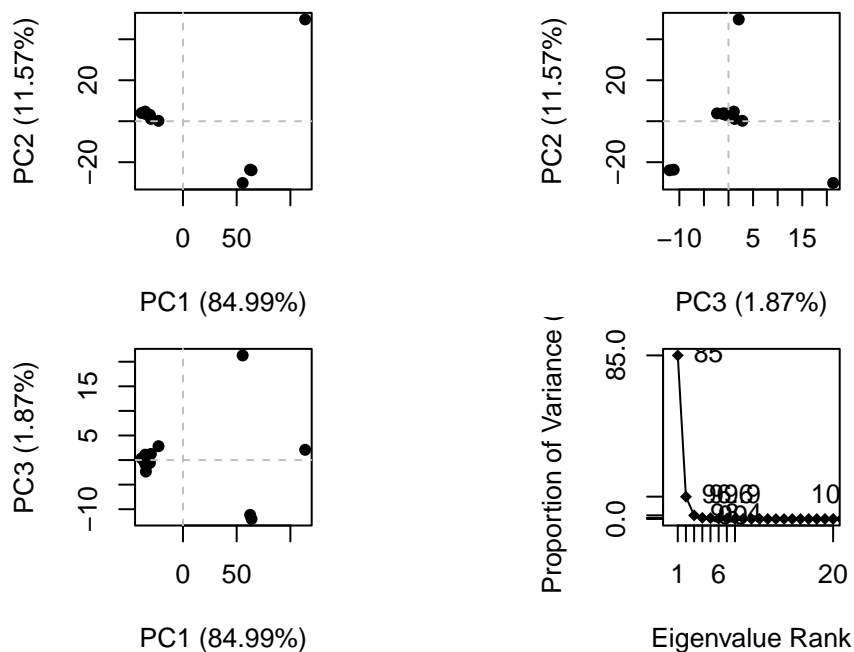
```
13 sequence rows; 227 position columns (204 non-gap, 23 gap)
```

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

Step 5. PCA

Let's use our old friend PCA to make sense of these confusing, complicated structure relationships.

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

showing there's 3 different clusters/ conformations

Let's make a trajectory (or movie) of the main conformational changes captured by PC1. We will use the `mktrj()` function for this...

```
mktrj(pc, file = "pca_result.pdb")
```

Upload file into molstar.org/viewer/ and see through playing video. Easier to view/ understand

Back of the Envelope Comparison of the PDB size to UniProt

```
uniprot <- 253206171
pdb <- 231029
pdb/uniprot *100
```

```
[1] 0.09124146
```