

# Comparative structure analysis

Thomas Bailey

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

```
library(bio3d)
```

**Step 1. retrieve 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  DELVIALVKERIAQEDCRNGFLLDGFRTIPQADAMKEAGINVDYVLEFDVPDELIVDRI  
      61      .      .      .      .      .      .      120  
  
      121     .      .      .      .      .      .      180  
pdb|1AKE|A  VGRRVHAPSGRVYHVKNPPKVEGKDDVTGEELTRKDDQEETVRKRLVEYHQMTPALIG  
      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. Blast search

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

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

```

Searching ... please wait (updates every 5 seconds) RID = V2NGYXZ5016
.
Reporting 87 hits

```

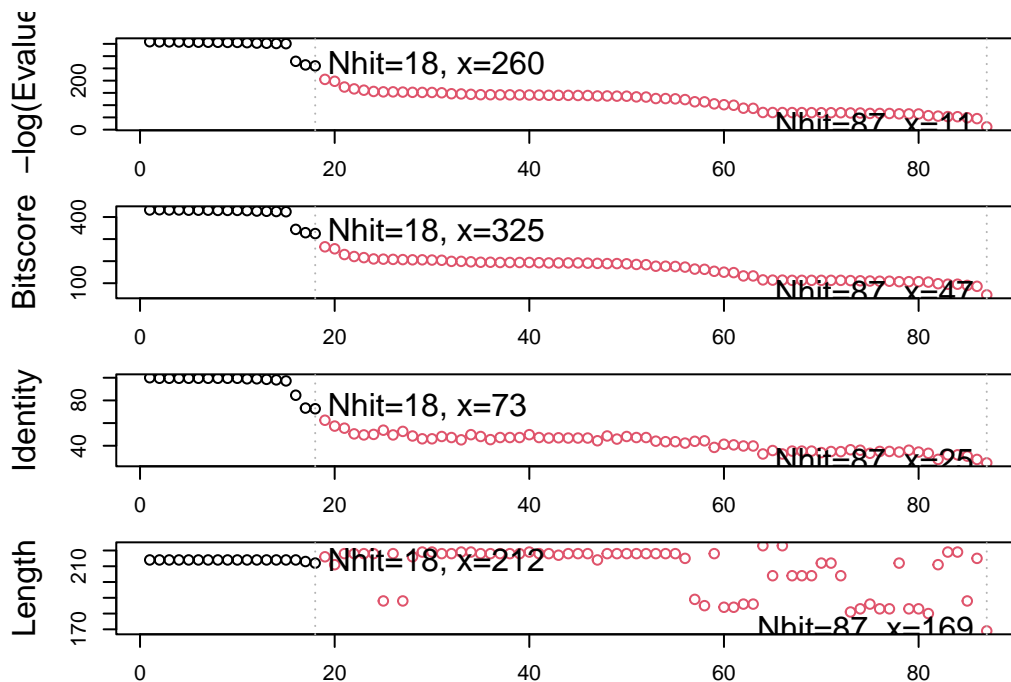
```
hits <- plot(blast)
```

```

* Possible cutoff values:    260 11
      Yielding Nhits:      18 87

* Chosen cutoff value of:    260
      Yielding Nhits:      18

```



### Step 3. Download all structures

These are our “top hits” i.e. all the structures in the PDB database related to our input sequence

```
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 exists. Skipping download
```

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

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

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

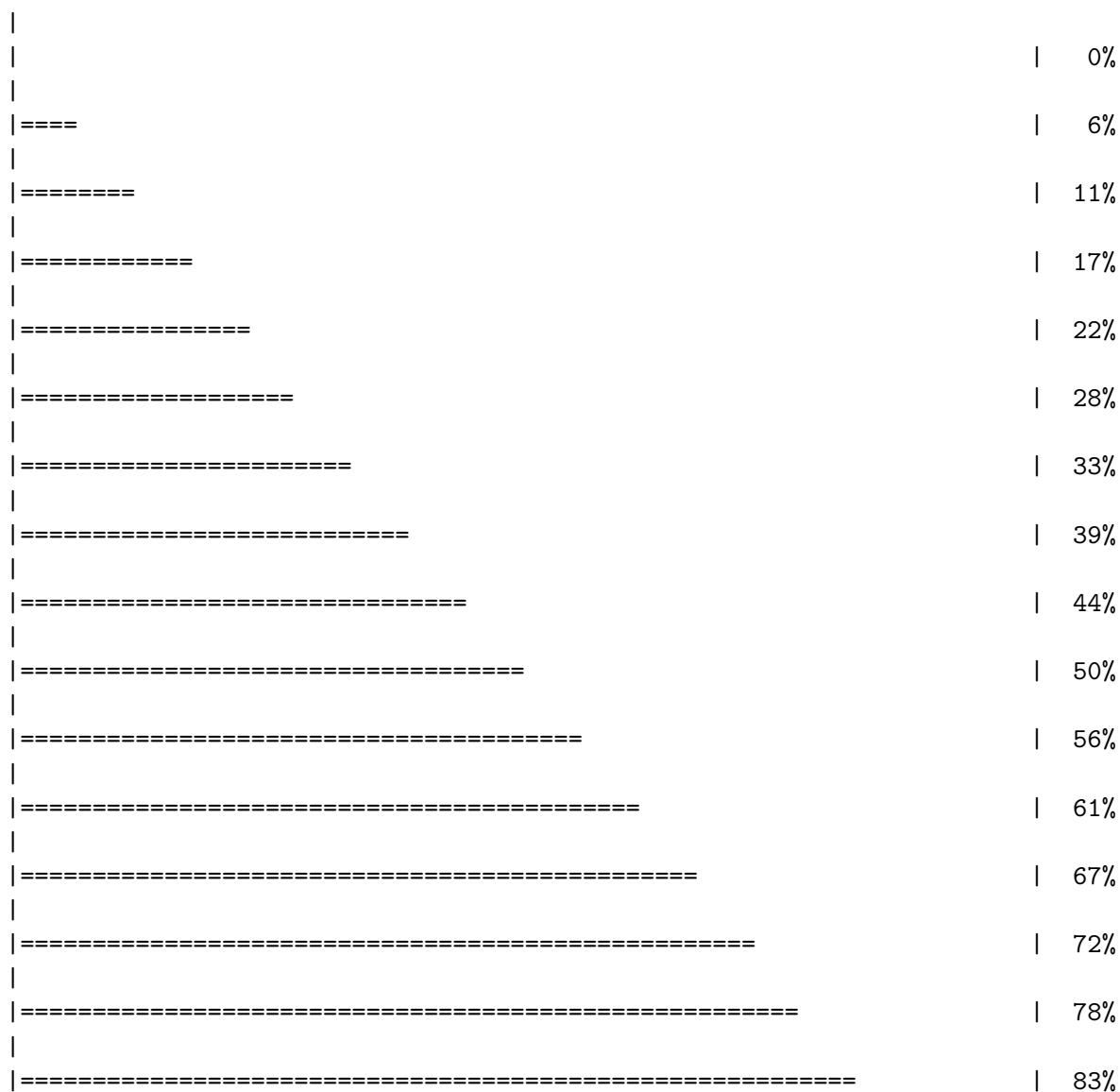
Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
pdbs/6HAP.pdb exists. Skipping download

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



```

|
|=====| 89%
|
|=====| 94%
|
|=====| 100%

```

#### step 4. Align and superpose

```
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
..
```

## Extracting sequences

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

## pdbc

	1	.	.	.	40
[Truncated_Name:1] 1AKE_A.pdb	--MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAA				
[Truncated_Name:2] 8BQF_A.pdb	--MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAA				
[Truncated_Name:3] 4X8M_A.pdb	--MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAA				
[Truncated_Name:4] 6S36_A.pdb	--MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAA				

```

[Truncated_Name:5]8Q2B_A.pdb      --MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMRLAA
[Truncated_Name:6]8RJ9_A.pdb      --MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMRLAA
[Truncated_Name:7]6RZE_A.pdb      --MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMRLAA
[Truncated_Name:8]4X8H_A.pdb      --MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMRLAA
[Truncated_Name:9]3HPR_A.pdb      --MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMRLAA
[Truncated_Name:10]1E4V_A.pdb     --MRIILLGAPVAGKGTQAQFIMEKYGIPQISTGDMRLAA
[Truncated_Name:11]5EJE_A.pdb     --MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMRLAA
[Truncated_Name:12]1E4Y_A.pdb     --MRIILLGALVAGKGTQAQFIMEKYGIPQISTGDMRLAA
[Truncated_Name:13]3X2S_A.pdb     --MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMRLAA
[Truncated_Name:14]6HAP_A.pdb     --MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMRLAA
[Truncated_Name:15]6HAM_A.pdb     --MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMRLAA
[Truncated_Name:16]8PVW_A.pdb     --MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMRLAA
[Truncated_Name:17]4K46_A.pdb     --MRIILLGAPGAGKGTQAQFIMAKFGIPQISTGDMRLAA
[Truncated_Name:18]4NP6_A.pdb     NAMRIILLGAPGAGKGTQAQFIMEKFGIPQISTGDMRLAA
                                *****  *~*****
1                                .                .                .                40

41                                .                .                .                80
[Truncated_Name:1]1AKE_A.pdb      VKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCR
[Truncated_Name:2]8BQF_A.pdb      VKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQE---
[Truncated_Name:3]4X8M_A.pdb      VKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCR
[Truncated_Name:4]6S36_A.pdb      VKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCR
[Truncated_Name:5]8Q2B_A.pdb      VKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCR
[Truncated_Name:6]8RJ9_A.pdb      VKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCR
[Truncated_Name:7]6RZE_A.pdb      VKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCR
[Truncated_Name:8]4X8H_A.pdb      VKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCR
[Truncated_Name:9]3HPR_A.pdb      VKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCR
[Truncated_Name:10]1E4V_A.pdb     VKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCR
[Truncated_Name:11]5EJE_A.pdb     VKSGSELGKQAKDIMDACKLVDELVIALVKERIAQEDCR
[Truncated_Name:12]1E4Y_A.pdb     VKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCR
[Truncated_Name:13]3X2S_A.pdb     VKSGSELGKQAKDIMDCGKLVDELVIALVKERIAQEDSR
[Truncated_Name:14]6HAP_A.pdb     VKSGSELGKQAKDIMDAGKLVDELVIALVRERICQEDSR
[Truncated_Name:15]6HAM_A.pdb     IKSGSELGKQAKDIMDAGKLVDEIIIALVKERICQEDSR
[Truncated_Name:16]8PVW_A.pdb     VKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCR
[Truncated_Name:17]4K46_A.pdb     IKAGTELGKQAKSVIDAGQLVSDDIILGLVKERIAQDDCA
[Truncated_Name:18]4NP6_A.pdb     IKAGTELGKQAKAVIDAGQLVSDDIILGLIKERIAQADCE
                                ~* *~*****  ^^*    **~*~^~^~*~^***  *
41                                .                .                .                80

81                                .                .                .                120
[Truncated_Name:1]1AKE_A.pdb      NGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVD
[Truncated_Name:2]8BQF_A.pdb      -GFLLDGFPR TIPQADAMKEAGINVDYVIEFDVPDELIVD
[Truncated_Name:3]4X8M_A.pdb      NGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVD

```



[Truncated_Name:4] 6S36_A.pdb	NGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVD
[Truncated_Name:5] 8Q2B_A.pdb	NGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVD
[Truncated_Name:6] 8RJ9_A.pdb	NGFLLAGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVD
[Truncated_Name:7] 6RZE_A.pdb	NGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVD
[Truncated_Name:8] 4X8H_A.pdb	NGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVD
[Truncated_Name:9] 3HPR_A.pdb	NGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVD
[Truncated_Name:10] 1E4V_A.pdb	NGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVD
[Truncated_Name:11] 5EJE_A.pdb	NGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVD
[Truncated_Name:12] 1E4Y_A.pdb	NGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVD
[Truncated_Name:13] 3X2S_A.pdb	NGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVD
[Truncated_Name:14] 6HAP_A.pdb	NGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVD
[Truncated_Name:15] 6HAM_A.pdb	NGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVD
[Truncated_Name:16] 8PVW_A.pdb	NGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVD
[Truncated_Name:17] 4K46_A.pdb	KGFLLDGFPR TIPQADGLKEVGVVVDYVIEFDVADSVIVE
[Truncated_Name:18] 4NP6_A.pdb	KGFLLDGFPR TIPQADGLKEMGINVDYVIEFDVADDVIVE
	**** *****^~** *^ ***** * ^**^
	81 . . . 120
	121 . . . 160
[Truncated_Name:1] 1AKE_A.pdb	RIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTGEELTTRKD
[Truncated_Name:2] 8BQF_A.pdb	RIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTGEELTTRKD
[Truncated_Name:3] 4X8M_A.pdb	RIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTGEELTTRKD
[Truncated_Name:4] 6S36_A.pdb	KIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTGEELTTRKD
[Truncated_Name:5] 8Q2B_A.pdb	RIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTGEELTTRKA
[Truncated_Name:6] 8RJ9_A.pdb	RIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTGEELTTRKD
[Truncated_Name:7] 6RZE_A.pdb	AIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTGEELTTRKD
[Truncated_Name:8] 4X8H_A.pdb	RIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTGEELTTRKD
[Truncated_Name:9] 3HPR_A.pdb	RIVGRRVHAPSGRVYHV KFNPPKVEGKDDGTGEELTTRKD
[Truncated_Name:10] 1E4V_A.pdb	RIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTGEELTTRKD
[Truncated_Name:11] 5EJE_A.pdb	RIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTGEELTTRKD
[Truncated_Name:12] 1E4Y_A.pdb	RIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTGEELTTRKD
[Truncated_Name:13] 3X2S_A.pdb	RIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTGEELTTRKD
[Truncated_Name:14] 6HAP_A.pdb	RIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTGEELTTRKD
[Truncated_Name:15] 6HAM_A.pdb	RIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTGEELTTRKD
[Truncated_Name:16] 8PVW_A.pdb	RILKRGE--TSGRV-----D
[Truncated_Name:17] 4K46_A.pdb	RMAGRRAHLASGR TYHNVNPPKVEGKDDVTGEDLVIRE
[Truncated_Name:18] 4NP6_A.pdb	RMAGRRAHLPSGR TYHVYNPPKVEGKDDVTGEDLVIRE
	^ * ***
	121 . . . 160
	161 . . . 200
[Truncated_Name:1] 1AKE_A.pdb	DQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDG
[Truncated_Name:2] 8BQF_A.pdb	DQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDG

[Truncated_Name:3] 4X8M_A.pdb	DQEETVRKRLVEWHQMTAPLIGYYSKEAEAGNTKYAKVDG	
[Truncated_Name:4] 6S36_A.pdb	DQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDG	
[Truncated_Name:5] 8Q2B_A.pdb	DQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDG	
[Truncated_Name:6] 8RJ9_A.pdb	DQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDG	
[Truncated_Name:7] 6RZE_A.pdb	DQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDG	
[Truncated_Name:8] 4X8H_A.pdb	DQEETVRKRLVEYHQMTAALIGYYSKEAEAGNTKYAKVDG	
[Truncated_Name:9] 3HPR_A.pdb	DQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDG	
[Truncated_Name:10] 1E4V_A.pdb	DQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDG	
[Truncated_Name:11] 5EJE_A.pdb	DQEECVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDG	
[Truncated_Name:12] 1E4Y_A.pdb	DQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDG	
[Truncated_Name:13] 3X2S_A.pdb	DQEETVRKRLCEYHQMTAPLIGYYSKEAEAGNTKYAKVDG	
[Truncated_Name:14] 6HAP_A.pdb	DQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDG	
[Truncated_Name:15] 6HAM_A.pdb	DQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDG	
[Truncated_Name:16] 8PVW_A.pdb	DNEETVRKRLVEYHQMTAPLIGYYSKEAEAGNTKYAKVDG	
[Truncated_Name:17] 4K46_A.pdb	DKEETVLARLGVYHNQTAPLIAYYGKEAEAGNTQYLKFDG	
[Truncated_Name:18] 4NP6_A.pdb	DKEETVRARLNVYHTQTAPLIEYYGKEAAAGKTQYLKFDG	
	* * * * *	
	161 . . . . . 200	
	201 . 216	
[Truncated_Name:1] 1AKE_A.pdb	TKPVAEVRADLEKILG	
[Truncated_Name:2] 8BQF_A.pdb	TKPVAEVRADLEKIL-	
[Truncated_Name:3] 4X8M_A.pdb	TKPVAEVRADLEKILG	
[Truncated_Name:4] 6S36_A.pdb	TKPVAEVRADLEKILG	
[Truncated_Name:5] 8Q2B_A.pdb	TKPVAEVRADLEKILG	
[Truncated_Name:6] 8RJ9_A.pdb	TKPVAEVRADLEKILG	
[Truncated_Name:7] 6RZE_A.pdb	TKPVAEVRADLEKILG	
[Truncated_Name:8] 4X8H_A.pdb	TKPVAEVRADLEKILG	
[Truncated_Name:9] 3HPR_A.pdb	TKPVAEVRADLEKILG	
[Truncated_Name:10] 1E4V_A.pdb	TKPVAEVRADLEKILG	
[Truncated_Name:11] 5EJE_A.pdb	TKPVAEVRADLEKILG	
[Truncated_Name:12] 1E4Y_A.pdb	TKPVAEVRADLEKILG	
[Truncated_Name:13] 3X2S_A.pdb	TKPVAEVRADLEKILG	
[Truncated_Name:14] 6HAP_A.pdb	TKPVCEVRADLEKILG	
[Truncated_Name:15] 6HAM_A.pdb	TKPVCEVRADLEKILG	
[Truncated_Name:16] 8PVW_A.pdb	TKPVAEVRADLEKILG	
[Truncated_Name:17] 4K46_A.pdb	TKAVAEVSAELEKALA	
[Truncated_Name:18] 4NP6_A.pdb	TKQVSEVSADIKALA	
	** * * * ^ * *	
	201 . 216	

Call:

```
pdbaln(files = files, fit = TRUE, exefile = "msa")
```

Class:

pdbs, fasta

Alignment dimensions:

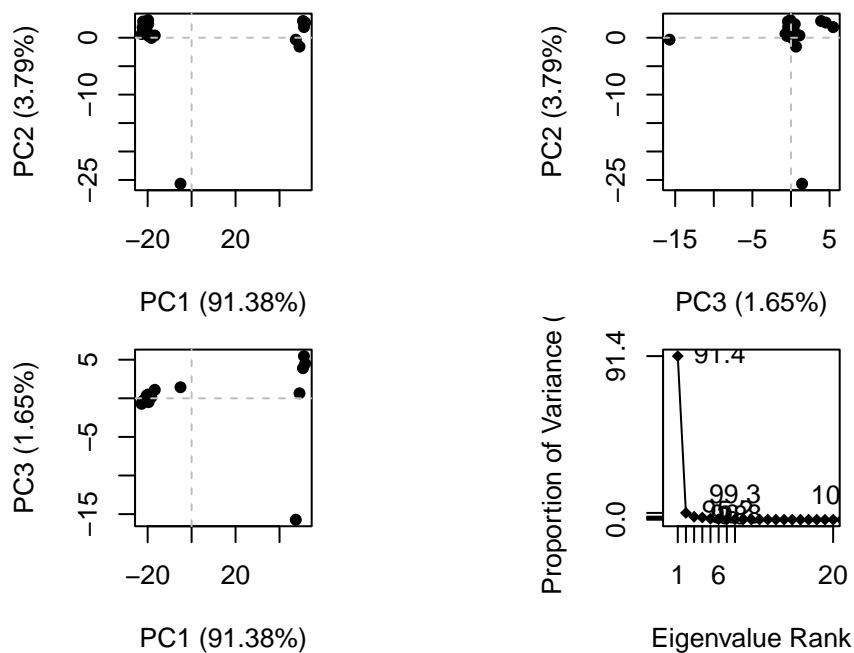
18 sequence rows; 216 position columns (182 non-gap, 34 gap)

+ 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 confussing, complicated structure relationships

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



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")
```

Back of the envelope comparison of the PDB size to UniProt

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

```
[1] 0.09124146
```