

# Class 11

Vanesa Fernandez

Today, before delving into structure prediction with AlfaFold we will finish off our previous lab10 “comparative structure analysis” section.

```
installed.packages("bio3d")
```

```
Package LibPath Version Priority Depends Imports LinkingTo Suggests
Enhances License License_is_FOSS License_restricts_use OS_type Archs
MD5sum NeedsCompilation Built
```

```
library(bio3d)
id <- "1ake_A"

aa <- get.seq(id)
```

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

Fetching... Please wait. Done.

```
print(aa)
```

```

      1      .      .      .      .      .      60
pdb|1AKE|A  MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDAGKLV
      1      .      .      .      .      .      60
      61      .      .      .      .      .      120
pdb|1AKE|A  DELVIALVKERIAQEDCRNGFLLDGFPRPTIPQADAMKEAGINVDYVLEFDVPDELIVDRI
      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

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

Searching ... please wait (updates every 5 seconds) RID = K4YXXZVB016

.....

Reporting 85 hits

```
attributes(b)
```

\$names

```
[1] "hit.tbl" "raw"      "url"
```

\$class

```
[1] "blast"
```

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

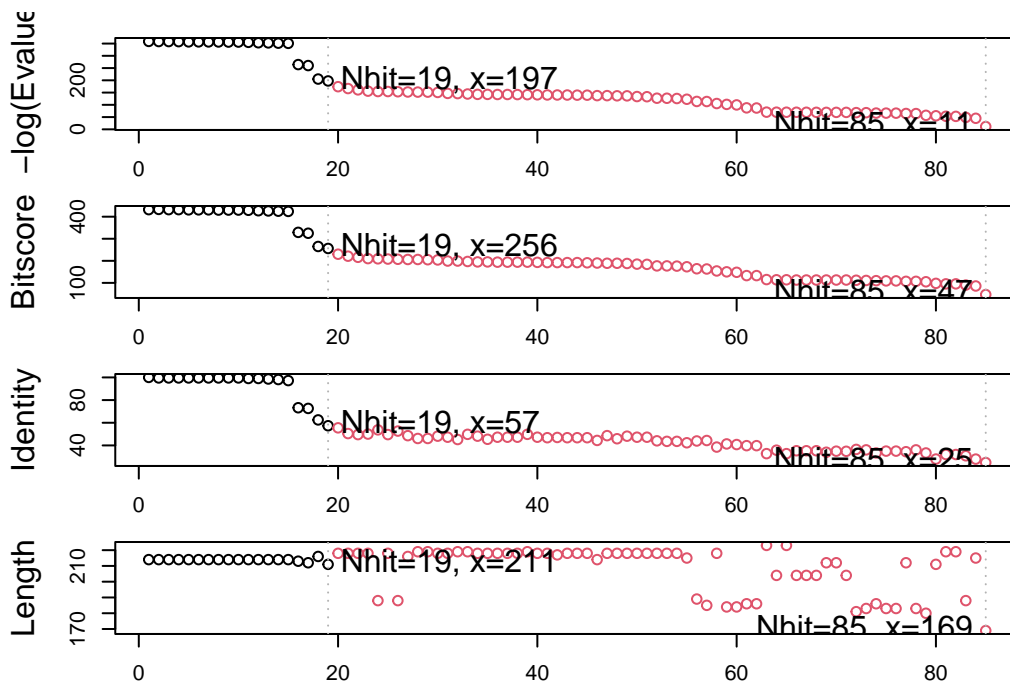
	queryid	subjectids	identity	alignmentlength	mismatches	gapopens	q.start
1	Query_1668479	1AKE_A	100.000	214	0	0	1
2	Query_1668479	8BQF_A	99.533	214	1	0	1
3	Query_1668479	4X8M_A	99.533	214	1	0	1
4	Query_1668479	6S36_A	99.533	214	1	0	1
5	Query_1668479	8Q2B_A	99.533	214	1	0	1

	6	Query_1668479	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	1.58e-156	432	100.00	358.7458	1AKE_A	1AKE_A
2	214	21	234	2.58e-156	433	100.00	358.2555	8BQF_A	8BQF_A
3	214	1	214	2.82e-156	432	100.00	358.1665	4X8M_A	4X8M_A
4	214	1	214	4.14e-156	432	100.00	357.7826	6S36_A	6S36_A
5	214	1	214	1.10e-155	431	99.53	356.8054	8Q2B_A	8Q2B_A
6	214	1	214	1.10e-155	431	99.53	356.8054	8RJ9_A	8RJ9_A

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

```
* Possible cutoff values: 197 11
    Yielding Nhits: 19 85

* Chosen cutoff value of: 197
    Yielding Nhits: 19
```



```
attributes(hits)
```

```
$names
[1] "hits" "pdb.id" "acc" "inds"
```

```
$class  
[1] "blast"
```

These are the top hits that we like from our blast results:

```
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" "4K46_A"  
[17] "4NP6_A" "3GMT_A" "4PZL_A"
```

```
# Download releated 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 = "pdbc", split = TRUE, gzip = TRUE):  
pdbc/8BQF.pdb.gz exists. Skipping download
```

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

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

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

```
Warning in get.pdb(hits$ pdb.id, path = "pdbc", split = TRUE, gzip = TRUE):  
pdbc/8RJ9.pdb.gz exists. Skipping download
```

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

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

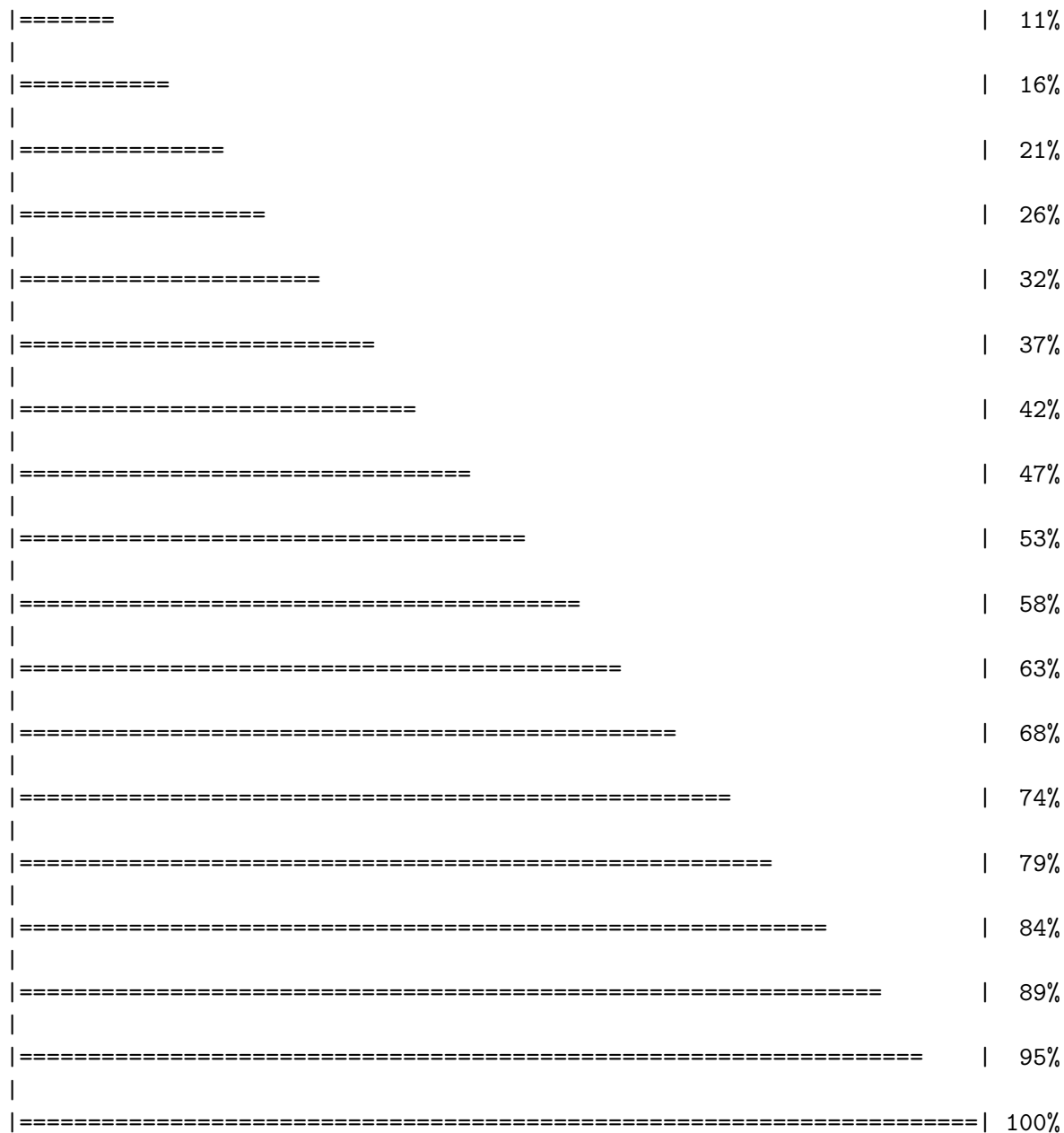
Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
pdbs/4K46.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
pdbs/4NP6.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
pdbs/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%
====	5%



I downloaded all ADK structures in the PDB database but viewing them is difficult as they need to be aligned and superimposed

To install BiocManager package from CRAN

```
##install.packages("BiocManager")
```

```
##BiocManager::install() ## <- use it to install any bioconductor package, BiocManager::inst
```

```
# Align related PDBs
```

```
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/4K46\_A.pdb

pdbbs/split\_chain/4NP6\_A.pdb

pdbbs/split\_chain/3GMT\_A.pdb

pdbbs/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

  ..   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/4K46\_A.pdb  
 PDB has ALT records, taking A only, rm.alt=TRUE  
 pdb/seq: 17 name: pdbs/split\_chain/4NP6\_A.pdb  
 pdb/seq: 18 name: pdbs/split\_chain/3GMT\_A.pdb  
 pdb/seq: 19 name: pdbs/split\_chain/4PZL\_A.pdb

## pdbs

	1	40
[Truncated_Name:1] 1AKE_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:2] 8BQF_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:3] 4X8M_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:4] 6S36_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:5] 8Q2B_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:6] 8RJ9_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:7] 6RZE_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:8] 4X8H_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:9] 3HPR_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS



[illegible]

[Truncated_Name:7] 6RZE_A.pdb		RIAQEDCRNGFLLDGFPR TIP QADAMKEAGINVDYVLEFD	
[Truncated_Name:8] 4X8H_A.pdb		RIAQEDCRNGFLLDGFPR TIP QADAMKEAGINVDYVLEFD	
[Truncated_Name:9] 3HPR_A.pdb		RIAQEDCRNGFLLDGFPR TIP QADAMKEAGINVDYVLEFD	
[Truncated_Name:10] 1E4V_A.pdb		RIAQEDCRNGFLLDGFPR TIP QADAMKEAGINVDYVLEFD	
[Truncated_Name:11] 5EJE_A.pdb		RIAQEDCRNGFLLDGFPR TIP QADAMKEAGINVDYVLEFD	
[Truncated_Name:12] 1E4Y_A.pdb		RIAQEDSRNGFLLDGFPR TIP QADAMKEAGINVDYVLEFD	
[Truncated_Name:13] 3X2S_A.pdb		RICQEDSRNGFLLDGFPR TIP QADAMKEAGINVDYVLEFD	
[Truncated_Name:14] 6HAP_A.pdb		RICQEDSRNGFLLDGFPR TIP QADAMKEAGINVDYVLEFD	
[Truncated_Name:15] 6HAM_A.pdb		RIAQDDCAKGFLLDGFPRTIPQADGLKEVGVVVDYVIEFD	
[Truncated_Name:16] 4K46_A.pdb		RIAQADCEKGFLLDGFPRTIPQADGLKEMGINVDYVIEFD	
[Truncated_Name:17] 4NP6_A.pdb		RLKEADCANGYLFDGFPRTIAQADAMKEAGVAIDYVLEID	
[Truncated_Name:18] 3GMT_A.pdb		RISKNDCCNNGFLLDGVPRTIPQAQLDKLGVNIDYIVEVD	
[Truncated_Name:19] 4PZL_A.pdb		*^                *~*     * **** ** ^      *^ ~**~*~ *	
	81	.	. . .
	121	.	. . .
[Truncated_Name:1] 1AKE_A.pdb		VPDELIVDRIVGRRVHAPS GRVYHV KFNPPKVEGKD DVTG	
[Truncated_Name:2] 8BQF_A.pdb		VPDELIVDRIVGRRVHAPS GRVYHV KFNPPKVEGKD DVTG	
[Truncated_Name:3] 4X8M_A.pdb		VPDELIVDRIVGRRVHAPS GRVYHV KFNPPKVEGKD DVTG	
[Truncated_Name:4] 6S36_A.pdb		VPDELIVDKI VGR RV H APS GRVYHV KFNPPKVEG KD D VT G	
[Truncated_Name:5] 8Q2B_A.pdb		VPDELIVDRIVGRRVHAPS GRVYHV KFNPPKVEGKD DVTG	
[Truncated_Name:6] 8RJ9_A.pdb		VPDELIVDRIVGRRVHAPS GRVYHV KFNPPKVEGKD DVTG	
[Truncated_Name:7] 6RZE_A.pdb		VPDELIVDA I VGR RV H APS GRVYHV KFNPPKVEG KD D VT G	
[Truncated_Name:8] 4X8H_A.pdb		VPDELIVDRIVGRRVHAPS GRVYHV KFNPPKVEGKD DVTG	
[Truncated_Name:9] 3HPR_A.pdb		VPDELIVDRIVGRRVHAPS GRVYHV KFNPPKVEGKD DG TG	
[Truncated_Name:10] 1E4V_A.pdb		VPDELIVDRIVGRRVHAPS GRVYHV KFNPPKVEGKD DVTG	
[Truncated_Name:11] 5EJE_A.pdb		VPDELIVDRIVGRRVHAPS GRVYHV KFNPPKVEGKD DVTG	
[Truncated_Name:12] 1E4Y_A.pdb		VPDELIVDRIVGRRVHAPS GRVYHV KFNPPKVEGKD DVTG	
[Truncated_Name:13] 3X2S_A.pdb		VPDELIVDRIVGRRVHAPS GRVYHV KFNPPKVEGKD DVTG	
[Truncated_Name:14] 6HAP_A.pdb		VPDELIVDRIVGRRVHAPS GRVYHV KFNPPKVEGKD DVTG	
[Truncated_Name:15] 6HAM_A.pdb		VADS VIVER MAG RRAHLASGR TYHN VN PP KV E GK DD VT G	
[Truncated_Name:16] 4K46_A.pdb		VADD VIVER MAG RRAH LPSGR TYHV VN PP KV E GK DD VT G	
[Truncated_Name:17] 4NP6_A.pdb		VP FSEIIERM SGR RH TP ASGR TY HV KF NP PK VE GK DD VT G	
[Truncated_Name:18] 3GMT_A.pdb		VAD N LLIERITGR RI HP ASGR TYHTKF NP PK VA DK DD VT G	
[Truncated_Name:19] 4PZL_A.pdb		*	
	121	.	. . .
	161	.	. . .
[Truncated_Name:1] 1AKE_A.pdb		EELTTRKDDQEETVRKR LV E Y HQ MT APLIGYS KEAE AGN	
[Truncated_Name:2] 8BQF_A.pdb		EELTTRKDDQEETVRKR LV E Y HQ MT APLIGYS KEAE AGN	
[Truncated Name:3] 4X8M_A.pdb		EELTTRKDDQEETVRKR LV E VH QMTA PLIGYS KEAE AGN	

[Truncated_Name:4] 6S36_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN	
[Truncated_Name:5] 8Q2B_A.pdb	EELTTRKADQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN	
[Truncated_Name:6] 8RJ9_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN	
[Truncated_Name:7] 6RZE_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN	
[Truncated_Name:8] 4X8H_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAALIGYYSKEAEAGN	
[Truncated_Name:9] 3HPR_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN	
[Truncated_Name:10] 1E4V_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN	
[Truncated_Name:11] 5EJE_A.pdb	EELTTRKDDQEECVRKRLVEYHQMTAPLIGYYSKEAEAGN	
[Truncated_Name:12] 1E4Y_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN	
[Truncated_Name:13] 3X2S_A.pdb	EELTTRKDDQEETVRKRLCEYHQMTAPLIGYYSKEAEAGN	
[Truncated_Name:14] 6HAP_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN	
[Truncated_Name:15] 6HAM_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN	
[Truncated_Name:16] 4K46_A.pdb	EDLVIREDDKEETVLARLGVYHNQTAPLIAYYGKEAEAGN	
[Truncated_Name:17] 4NP6_A.pdb	EDLVIREDDKEETVRARLNVYHTQTAPLIEYYGKEAAAAGK	
[Truncated_Name:18] 3GMT_A.pdb	EPLVQRDDKEETVKKRLDVYEAQTKPLITYYGDWARRGA	
[Truncated_Name:19] 4PZL_A.pdb	EPLITRTDDNEDTVKQRLSVYHAQTAKLIDFYRNFSSNTNT	
	* * * * ^ * ** ^ * ** ^ *	
	161 . . . 200	
	201 . . 227	
[Truncated_Name:1] 1AKE_A.pdb	T--KYAKVDGTKPVAEVRADLEKILG-	
[Truncated_Name:2] 8BQF_A.pdb	T--KYAKVDGTKPVAEVRADLEKIL--	
[Truncated_Name:3] 4X8M_A.pdb	T--KYAKVDGTKPVAEVRADLEKILG-	
[Truncated_Name:4] 6S36_A.pdb	T--KYAKVDGTKPVAEVRADLEKILG-	
[Truncated_Name:5] 8Q2B_A.pdb	T--KYAKVDGTKPVAEVRADLEKILG-	
[Truncated_Name:6] 8RJ9_A.pdb	T--KYAKVDGTKPVAEVRADLEKILG-	
[Truncated_Name:7] 6RZE_A.pdb	T--KYAKVDGTKPVAEVRADLEKILG-	
[Truncated_Name:8] 4X8H_A.pdb	T--KYAKVDGTKPVAEVRADLEKILG-	
[Truncated_Name:9] 3HPR_A.pdb	T--KYAKVDGTKPVAEVRADLEKILG-	
[Truncated_Name:10] 1E4V_A.pdb	T--KYAKVDGTKPVAEVRADLEKILG-	
[Truncated_Name:11] 5EJE_A.pdb	T--KYAKVDGTKPVAEVRADLEKILG-	
[Truncated_Name:12] 1E4Y_A.pdb	T--KYAKVDGTKPVAEVRADLEKILG-	
[Truncated_Name:13] 3X2S_A.pdb	T--KYAKVDGTKPVAEVRADLEKILG-	
[Truncated_Name:14] 6HAP_A.pdb	T--KYAKVDGTKPVCEVRADLEKILG-	
[Truncated_Name:15] 6HAM_A.pdb	T--KYAKVDGTKPVCEVRADLEKILG-	
[Truncated_Name:16] 4K46_A.pdb	T--QYLKFDGTKA VAESAELEKALA-	
[Truncated_Name:17] 4NP6_A.pdb	T--QYLKFDGTKQVSEVSADIAKALA-	
[Truncated_Name:18] 3GMT_A.pdb	E-----NGLKAPA-----YRKISG-	
[Truncated_Name:19] 4PZL_A.pdb	KIPKYIKINGDQAVEKVSQDIFDQLNK	
	*	
	201 . . 227	

Call:

```

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

Class:
  pdbs, fasta

Alignment dimensions:
  19 sequence rows; 227 position columns (199 non-gap, 28 gap)

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

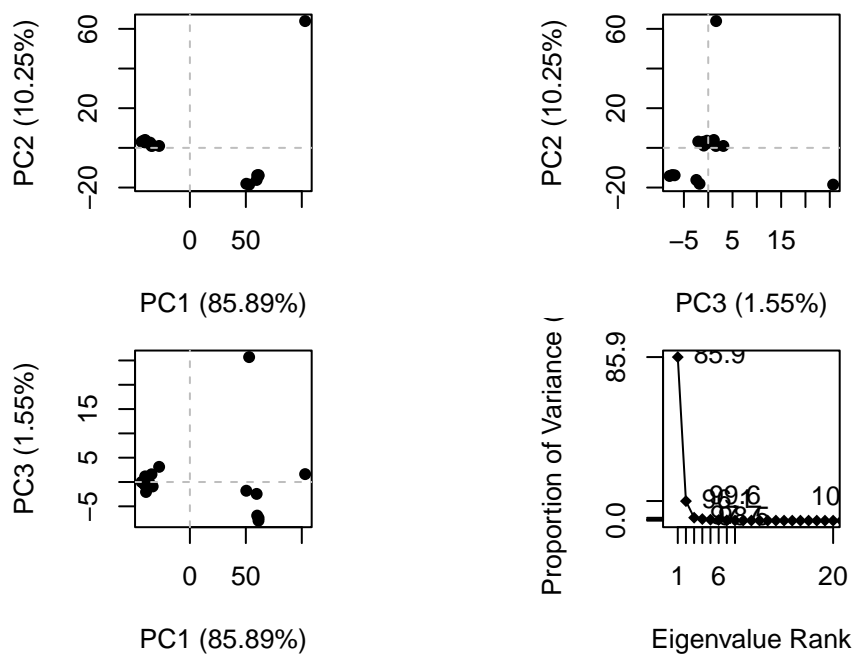
```

## Principal component analysis

```

# Perform Principal Component Analysis on the aligned PDB structures
pc <- pca(pdbs)
plot(pc)

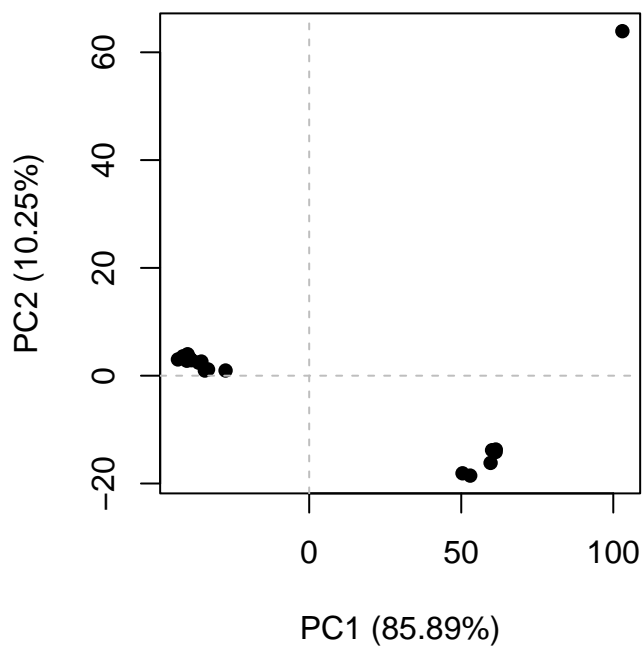
```



```

# Plot the first two principal components (PC1 vs PC2)
plot(pc, pc.axes = c(1,2))

```



**To examine in more detail what PC1 (or any PC) is capturing here we can plot the loadings or make a wee movie (trajectory) of moving along PC1**

#note: the zero is the average... the more you move in any direction you do away from the average towards whatever the PC is attributed to.

```
mktrj(pc, pc=1, file="pc1.pdb")
```

##Visualize the molecule structure in mol\* <https://molstar.org/viewer/> #1) Open this file in mol\* #2) Start animation #3) Export animation if need be

##This is the first PC, and doing it this way is much easier than doing it one by one