

Class 10: Structural Bioninformatics II

Moises Gonzalez (A17579866)

Q10. Which of the packages above is found only on BioConductor and not CRAN?

msa

Q11. Which of the above packages is not found on BioConductor or CRAN?:

bio3d-view

Q12. True or False? Functions from the devtools package can be used to install packages from GitHub and BitBucket?

TRUE

```
library(bio3d)
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  MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGMDLRAAVKSGSELGKQAKDIMDAGKLV
      1      .      .      .      .      .      60

      61      .      .      .      .      .      120
pdb|1AKE|A  DELVIALVKERIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVDRI
      61      .      .      .      .      .      120

     121      .      .      .      .      .      180
pdb|1AKE|A  VGRRVHAPSGRVYHV KFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
```


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/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

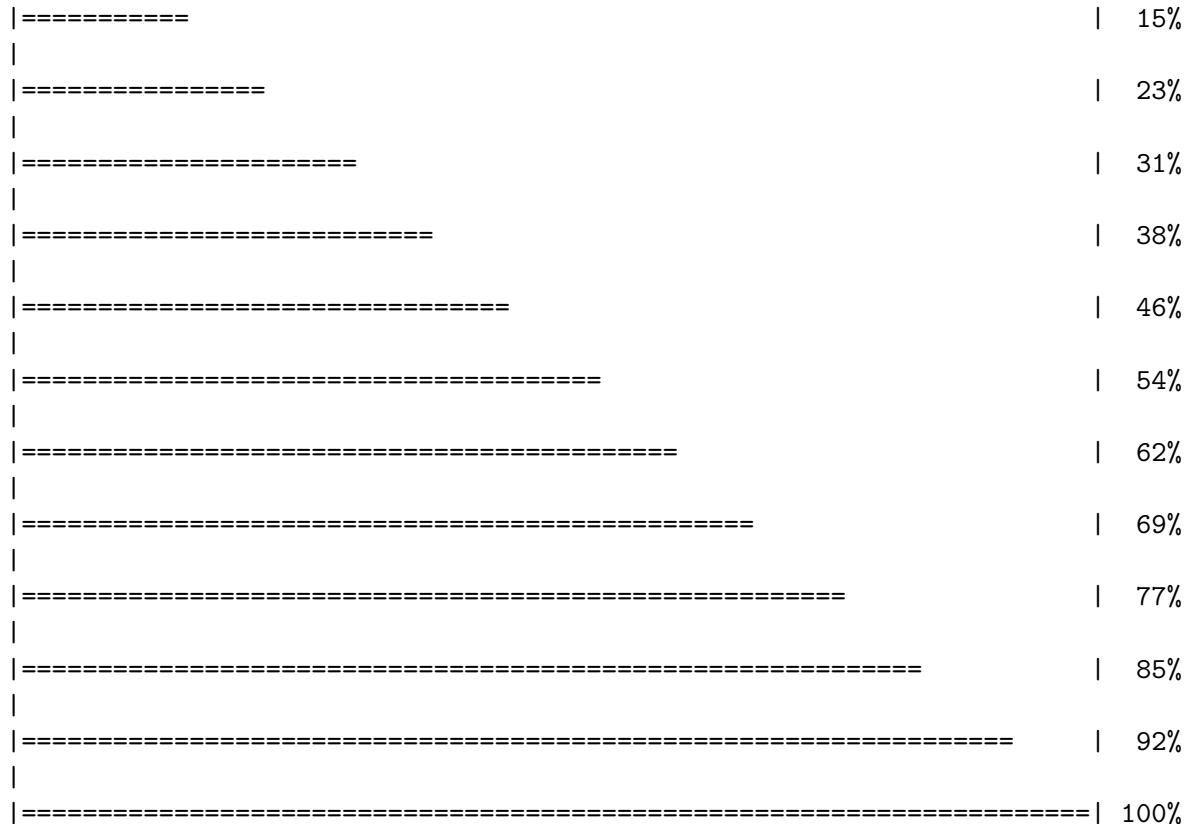
Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/6HAM.pdb.gz exists. Skipping download

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/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%



Now I want to align and superpose these structures which are all over the place.

```
# Align related PDBs
pdbc <- pdbcaln(files, fit = TRUE, exefile="msa")
```

Reading PDB files:

```
pdbc/split_chain/1AKE_A.pdb
pdbc/split_chain/6S36_A.pdb
pdbc/split_chain/6RZE_A.pdb
pdbc/split_chain/3HPR_A.pdb
pdbc/split_chain/1E4V_A.pdb
pdbc/split_chain/5EJE_A.pdb
pdbc/split_chain/1E4Y_A.pdb
pdbc/split_chain/3X2S_A.pdb
pdbc/split_chain/6HAP_A.pdb
pdbc/split_chain/6HAM_A.pdb
pdbc/split_chain/4K46_A.pdb
pdbc/split_chain/3GMT_A.pdb
```

```

pdb/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: pdb/split_chain/1AKE_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 2   name: pdb/split_chain/6S36_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 3   name: pdb/split_chain/6RZE_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 4   name: pdb/split_chain/3HPR_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 5   name: pdb/split_chain/1E4V_A.pdb
pdb/seq: 6   name: pdb/split_chain/5EJE_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 7   name: pdb/split_chain/1E4Y_A.pdb
pdb/seq: 8   name: pdb/split_chain/3X2S_A.pdb
pdb/seq: 9   name: pdb/split_chain/6HAP_A.pdb
pdb/seq: 10  name: pdb/split_chain/6HAM_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 11  name: pdb/split_chain/4K46_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 12  name: pdb/split_chain/3GMT_A.pdb
pdb/seq: 13  name: pdb/split_chain/4PZL_A.pdb

```

Let's have a look at our `pdb` object:

```
pdb
```

	1	.	.	.	40
[Truncated_Name:1] 1AKE_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS			
[Truncated_Name:2] 6S36_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS			
[Truncated_Name:3] 6RZE_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS			
[Truncated_Name:4] 3HPR_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS			

[Truncated_Name:5] 1E4V_A.pdb	-----MRIILLGAPVAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:6] 5EJE_A.pdb	-----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:7] 1E4Y_A.pdb	-----MRIILLGALVAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:8] 3X2S_A.pdb	-----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:9] 6HAP_A.pdb	-----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:10] 6HAM_A.pdb	-----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:11] 4K46_A.pdb	-----MRIILLGAPGAGKGTQAQFIMAKFGIPQIS
[Truncated_Name:12] 3GMT_A.pdb	-----MRLILLGAPGAGKGTQANFIKEKFIPQIS
[Truncated_Name:13] 4PZL_A.pdb	TENLYFQSNAMRIILLGAPGAGKGTQAKIIEQKNIAHIS
	^*** ***** * ^ * **
1	. . . 40
	41 . . . 80
[Truncated_Name:1] 1AKE_A.pdb	TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:2] 6S36_A.pdb	TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:3] 6RZE_A.pdb	TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:4] 3HPR_A.pdb	TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:5] 1E4V_A.pdb	TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:6] 5EJE_A.pdb	TGDMLRAAVKSGSELGKQAKDIMACKLVTDDELVIALVKE
[Truncated_Name:7] 1E4Y_A.pdb	TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:8] 3X2S_A.pdb	TGDMLRAAVKSGSELGKQAKDIMDCGLVTDELVIALVKE
[Truncated_Name:9] 6HAP_A.pdb	TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVRE
[Truncated_Name:10] 6HAM_A.pdb	TGDMLRAAIKSGSELGKQAKDIMDAGKLVTDEIIIALVKE
[Truncated_Name:11] 4K46_A.pdb	TGDMLRAAIKAGTELGKQAKSVIDAGQLVSDDIILGLVKE
[Truncated_Name:12] 3GMT_A.pdb	TGDMLRAAVKAGTPLGVEAKTYMDEGKLPDSLIIIGLVKE
[Truncated_Name:13] 4PZL_A.pdb	TGDMIRETIKSGSALGQELKKVLDA GELVSDEFIIKIVKD
	****~* ~* *~** * ~* ** * ^^ ~*^^
41	. . . 80
	81 . . . 120
[Truncated_Name:1] 1AKE_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:2] 6S36_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:3] 6RZE_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:4] 3HPR_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:5] 1E4V_A.pdb	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	RIAQQDDCAKGFLLDGFPR TIPQADGLKEVGVVVDYVIEFD
[Truncated_Name:12] 3GMT_A.pdb	RLKEADCANGYLDFGFPRTIAQADAMKEAGVAIDYVLEID
[Truncated_Name:13] 4PZL_A.pdb	RISKNDCNNGFLLDGVPR TIPQAQELD KLGVNIDYIVEVD

```

      *~      *      *~* ** ***** **      ^      *~ ^**~*~* *
81          .          .          .          120

121          .          .          .          160
[Truncated_Name:1] 1AKE_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:2] 6S36_A.pdb      VPDELIVDKIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:3] 6RZE_A.pdb      VPDELIVDAIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:4] 3HPR_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDGTG
[Truncated_Name:5] 1E4V_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:6] 5EJE_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:7] 1E4Y_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:8] 3X2S_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:9] 6HAP_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:10] 6HAM_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG
[Truncated_Name:11] 4K46_A.pdb      VADSVIVERMAGRRAHLASGRTYHNVNPPKVEGKDDVTG
[Truncated_Name:12] 3GMT_A.pdb      VPFSEIIERMSGRRTHPASGRTYHVKNPPKVEGKDDVTG
[Truncated_Name:13] 4PZL_A.pdb      VADNLLIERITGRRIH PASGRTYHTKFNPPKVADKDDVTG
      *      ^^^ ^ *** *      *** **      ^*****      *** **
121          .          .          .          160

161          .          .          .          200
[Truncated_Name:1] 1AKE_A.pdb      EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:2] 6S36_A.pdb      EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:3] 6RZE_A.pdb      EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:4] 3HPR_A.pdb      EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:5] 1E4V_A.pdb      EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:6] 5EJE_A.pdb      EELTTRKDDQEECVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:7] 1E4Y_A.pdb      EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:8] 3X2S_A.pdb      EELTTRKDDQEETVRKRLCEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:9] 6HAP_A.pdb      EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:10] 6HAM_A.pdb      EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:11] 4K46_A.pdb      EDLVIREDDKEETV LARLG VYHNQTAPLIAYYGKEAEAGN
[Truncated_Name:12] 3GMT_A.pdb      EPLVQRDDDEETVKKRLDVYEAQTKPLITYYGDWARRGA
[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--QYLKFDGTKAVAEVSAELEKALA-
[Truncated_Name:12]3GMT_A.pdb     E-----NGLKAPA-----YRKISG-
[Truncated_Name:13]4PZL_A.pdb     KIPKYIKINGDQAVEKVSQDIFDQLNK
                                   *
                                   .           .           227
201

```

Call:

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

Class:

```
pdbs, fasta
```

Alignment dimensions:

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

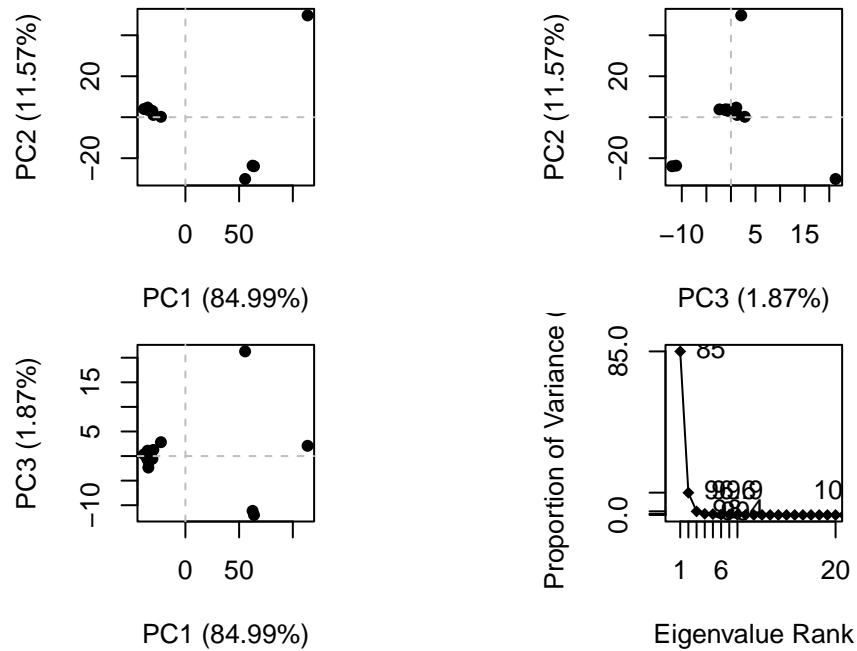
```
+ attr: xyz, resno, b, chain, id, ali, resid, sse, call
```

Now we have our aligned and superposed structures we can perform all sorts of analysis on them. Let's do PCA

```

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

```

Results of PCA on Adenylate kinase X-ray structures. Each dot represents one PDB structure.

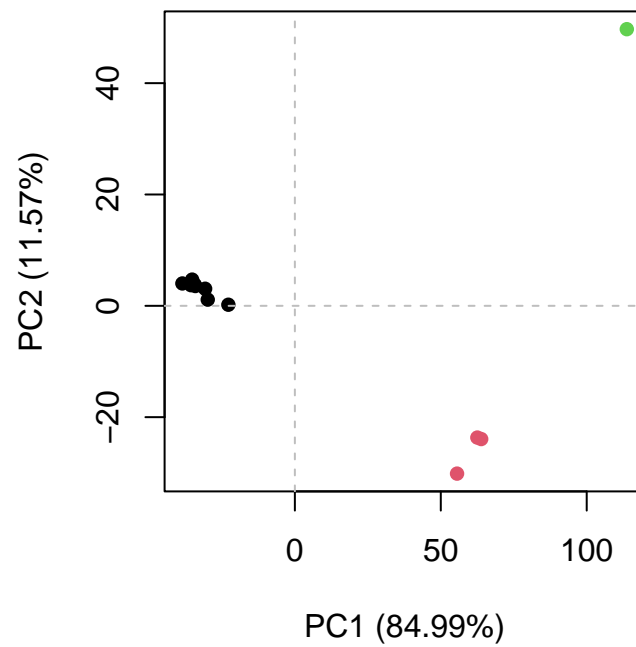
We can cluster the structures by RMSD (or any other method).

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

Warning in rmsd(pdb): No indices provided, using the 204 non NA positions

```
# Structure-based clustering
hc.rd <- hclust(dist(rd))
grps <- cutree(hc.rd, k=3)

plot(pc.xray, 1:2, col=grps)
```



We can make a wee movie- also called a trajectory of the major differences (i.e. structural displacements) of ADK.

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