

Class 10: Structural Bioinformatics Part 2

Peter Shamasha (A15857589)

Structural Bioinformatics Part 2

Comparative analysis of ADK

ADK (Adenylate kinase) is an important enzyme that maintains the equilibrium between cytoplasmic nucleotides essential for many cellular processes.

We will perform principal component analysis (PCA) on the complete collection of Adenylate kinase structures in the protein data-bank (PDB).

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?:

Grantlab/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  MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLV
      1      .      .      .      .      .      .      60

      61      .      .      .      .      .      .      120
pdb|1AKE|A  DELVIALVKERIAQEDCRNGFLLDGFPRTPQADAMKEAGINVDYVLEFDVPDELIVDRI
      61      .      .      .      .      .      .      120

      121      .      .      .      .      .      .      180
pdb|1AKE|A  VGRRVHAPSGRVYHVKNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTPALIG
      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
```

Q13. How many amino acids are in this sequence, i.e. how long is this sequence?

214

```

# Blast
#b <- blast.pdb(aa)

# Plot a summary of search results
#hits <- plot(b)

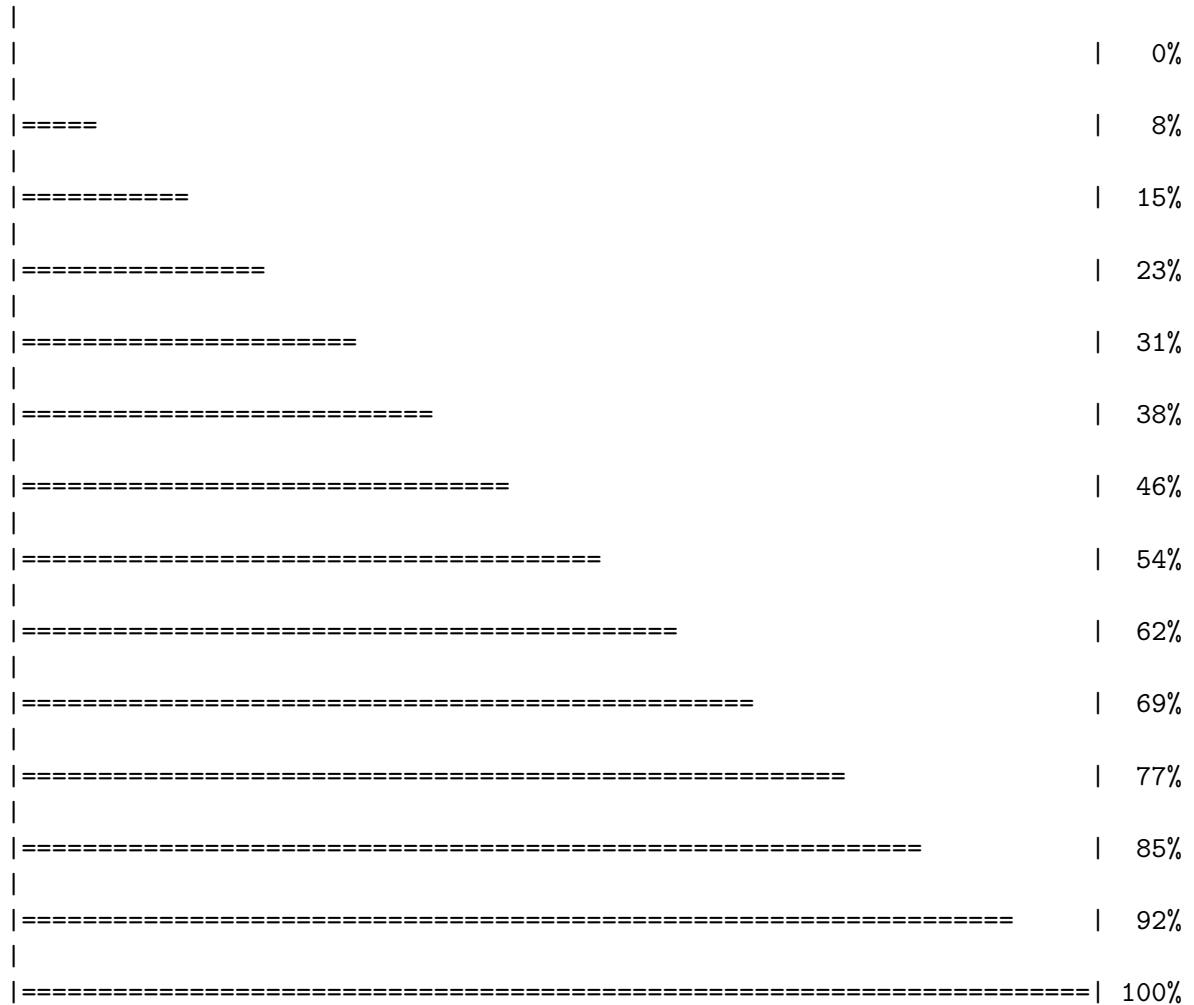
#List out some "top hits"
#head(hits$ pdb.id)

```



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

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



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

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

Reading PDB files:

pdb/split_chain/1AKE_A.pdb
pdb/split_chain/6S36_A.pdb
pdb/split_chain/6RZE_A.pdb
pdb/split_chain/3HPR_A.pdb
pdb/split_chain/1E4V_A.pdb
pdb/split_chain/5EJE_A.pdb
pdb/split_chain/1E4Y_A.pdb
pdb/split_chain/3X2S_A.pdb
pdb/split_chain/6HAP_A.pdb
pdb/split_chain/6HAM_A.pdb
pdb/split_chain/4K46_A.pdb
pdb/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: pdbs/split_chain/3GMT_A.pdb
pdb/seq: 13   name: pdbs/split_chain/4PZL_A.pdb

```

Let's Have a look at our pdbs object:

```
pdbs
```

```

[Truncated_Name:1] 1AKE_A.pdb      1          .          .          .          40
[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      -----MRLILLGAPGAGKGTQANFIKEKFGIPQIS
[Truncated_Name:13] 4PZL_A.pdb      TENLYFQSNAMRIILLGAPGAGKGTQAKIIEQKYNIAHIS
                                   **~*****  *****  *  *~ *  **
1          .          .          .          40

41          .          .          .          80
[Truncated_Name:1] 1AKE_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE
[Truncated_Name:2] 6S36_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE
[Truncated_Name:3] 6RZE_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE
[Truncated_Name:4] 3HPR_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE
[Truncated_Name:5] 1E4V_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE
[Truncated_Name:6] 5EJE_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDACKLVTDLVIALVKE
[Truncated_Name:7] 1E4Y_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE
[Truncated_Name:8] 3X2S_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDCGKLVTDLVIALVKE
[Truncated_Name:9] 6HAP_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVRE
[Truncated_Name:10] 6HAM_A.pdb      TGDMLRAAIKSGSELGKQAKDIMDAGKLVTDLIIIALVKE
[Truncated_Name:11] 4K46_A.pdb      TGDMLRAAIKAGTELGKQAKSVIDAGQLVSDDIILGLVKE
[Truncated_Name:12] 3GMT_A.pdb      TGDMLRAAVKAGTPLGVEAKTYMDEGKLVPSLIIGLVKE
[Truncated_Name:13] 4PZL_A.pdb      TGDMIRETIKSGSALGQELKKVLDAGELVSDEFIIVKIVKD
****~*  ~*  *~ **  *  ~*  ** *  ^^  ~*^^
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	RIAQDDCAKGFLDGFPR TIPQADGLKEVG VVDYVIEFD
[Truncated_Name:12] 3GMT_A.pdb	RLKEADCANGYLF DGFPR TIAQADAMKEAGVAIDYVLEID
[Truncated_Name:13] 4PZL_A.pdb	RISKNCNNGFLLDGVPR 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 TYHNVNPPKVEGKDDVTG
[Truncated_Name:12] 3GMT_A.pdb	VPFSEIIERMSGRRTHPASGR TYHV KFNPPKVEGKDDVTG
[Truncated_Name:13] 4PZL_A.pdb	VADNLLIERITGRR IHPASGR TYHTKFNPPKVADKDDVTG
	* ^^^ ^ *** * *** ** ^***** *** **
	121 . . . 160
	161 . . . 200
[Truncated_Name:1] 1AKE_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAP LIGYYSKEAEAGN
[Truncated_Name:2] 6S36_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAP LIGYYSKEAEAGN
[Truncated_Name:3] 6RZE_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAP LIGYYSKEAEAGN
[Truncated_Name:4] 3HPR_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAP LIGYYSKEAEAGN
[Truncated_Name:5] 1E4V_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAP LIGYYSKEAEAGN
[Truncated_Name:6] 5EJE_A.pdb	EELTTRKDDQEECVRKRLVEYHQMTAP LIGYYSKEAEAGN
[Truncated_Name:7] 1E4Y_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAP LIGYYSKEAEAGN
[Truncated_Name:8] 3X2S_A.pdb	EELTTRKDDQEETVRKRLCEYHQMTAP LIGYYSKEAEAGN
[Truncated_Name:9] 6HAP_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAP LIGYYSKEAEAGN

```

[Truncated_Name:10] 6HAM_A.pdb    EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:11] 4K46_A.pdb    EDLVIREDDKEETVLARLGVYHNQTAPLIAYYGKEAEAGN
[Truncated_Name:12] 3GMT_A.pdb    EPLVQRDDDKEETVKKRLDVYEAQTKPLITYYGDWARRGA
[Truncated_Name:13] 4PZL_A.pdb    EPLITRTDDNEDTVKQRLSVYHAQTAKLIDFYRNFSSNT
                                   * * * ** *~ * ** * * ** ~*
161                               .           .           .           200

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

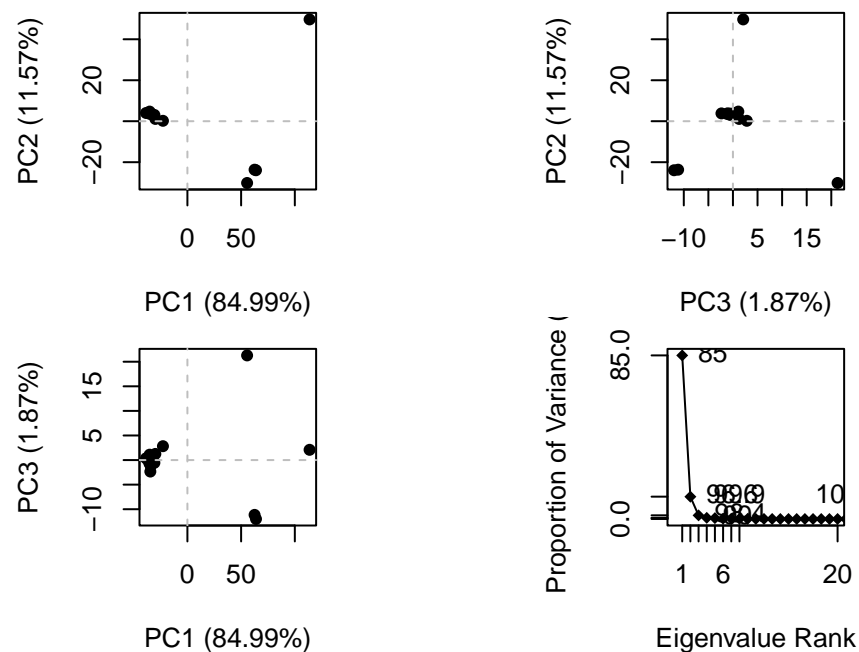
```
+ attr: xyz, resno, b, chain, id, ali, resid, sse, call
```

Now we have our aligned and superimposed 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.

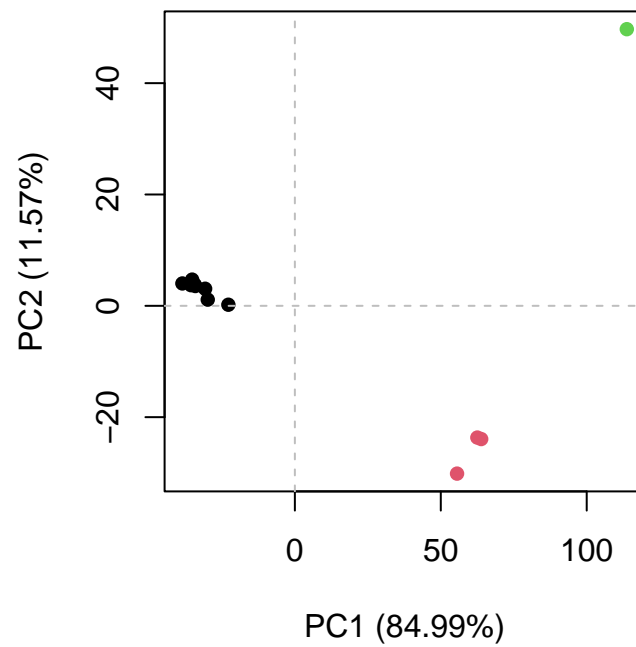
We can cluster the structures by RMSD (or any other method).

```
rd <- rmsd(pdbbs)
```

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

```
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")
```